Maria Sjöholm

International Human Rights Law and Protection Against Gender-Based Harm on the Internet

Springer

International Human Rights Law and Protection Against Gender-Based Harm on the Internet

Maria Sjöholm

International Human Rights Law and Protection Against Gender-Based Harm on the Internet

Springer

Maria Sjöholm
JPS
Örebro University
Örebro, Sweden

ISBN 978-3-031-15865-0 ISBN 978-3-031-15866-7 (eBook)
https://doi.org/10.1007/978-3-031-15866-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgments

The work on this monograph has been made possible through the support of multiple sources, for which I am sincerely grateful. Torsten Söderbergs Stiftelse provided a research grant for this project during a period of 2 years (2016–2018). As the project expanded, Stiftelsen för Rättsvetenskaplig Forskning contributed with financial support for an additional research period in the spring of 2020. Längmanska Kulturfonden awarded a grant for language editing by Louise Ratford. The grant by Torsten Söderbergs Stiftelse also allowed for a research visit with Prof. Donna M. Hughes of the University of Rhode Island, who generously assisted me with valuable advice and resources relevant for the book. I also wish to thank my colleagues at Örebro University and my family for their support.

I have sought to state the law as it stood on 15 March 2022.

Örebro University, Örebro, Sweden — Maria Sjöholm
April 2022

Contents

1 Introduction 1
1.1 Background 1
1.2 Research Aims and Questions 5
1.3 Methods, Theories and Materials 7
1.4 Delimitations 11
References 13

2 The Internet: A Gendered Space 15
2.1 Introduction 15
2.2 A Rights-Based Approach to the Internet 16
2.2.1 The Role for International Human Rights Law in Internet Regulation 16
2.2.2 A Gender Equal Internet 20
2.3 Gendering Features of the Internet 54
2.3.1 Introduction 54
2.3.2 Constraints of User Behaviour 56
2.4 Conclusion 68
References 70

3 Challenges in International Human Rights Law 75
3.1 Introduction 75
3.2 What Is Harmful? 76
3.2.1 Introduction 76
3.2.2 Technosocial Harm 77
3.2.3 Theories on Harm 82
3.3 The Scope of Rights Online and Offline: Harm, Values and Concepts 101
3.3.1 Introduction 101
3.3.2 The Freedom of Expression 102
3.3.3 The Right to Privacy 122

3.3.4 Proportionality Assessments and Balancing in Conflicts of Rights . . . 138
3.3.5 Summary . . . 142
3.4 Who Is Liable? . . . 143
3.4.1 Introduction . . . 143
3.4.2 Individual Perpetrators and User Anonymity . . . 144
3.4.3 Liability of Internet Intermediaries and Media Publishers . . . 151
3.4.4 Conclusion . . . 190
References . . . 195

4 Online Gender-Based Offences and International Human Rights Law . . . 203
4.1 Introduction . . . 203
4.2 Sexual Violence . . . 204
4.2.1 Introduction . . . 204
4.2.2 The Prohibition and Definition of Sexual Violence . . . 205
4.2.3 Hierarchies of Sexual Violence . . . 211
4.2.4 Obligations to Protect . . . 216
4.2.5 Conclusions . . . 230
4.3 Harassment . . . 234
4.3.1 Introduction . . . 234
4.3.2 Sexual Harassment . . . 236
4.3.3 Threats of Violence and Disclosure . . . 250
4.3.4 Defamation . . . 255
4.3.5 The Disclosure of Private Information . . . 271
4.4 Hate Speech . . . 284
4.4.1 Theorising Harm . . . 284
4.4.2 International Human Rights Law . . . 288
4.4.3 Online Hate Speech . . . 300
4.4.4 Sexist Hate Speech . . . 303
4.5 Harmful Pornography . . . 314
4.5.1 Introduction . . . 314
4.5.2 Theorising Harm . . . 315
4.5.3 Online Pornography . . . 321
4.5.4 International Human Rights Law . . . 324
4.5.5 Conclusion . . . 334
References . . . 336

5 Summary and Conclusion . . . 343
5.1 Gendered Spheres . . . 343
5.2 The Scope of Rights: Values, Harm and Balancing . . . 346
5.3 Obligations and Liability . . . 351
Reference . . . 356

Index . . . 357

Abbreviations and Acronyms

AI	Artificial Intelligence
ACHR	American Convention on Human Rights 1969
ACmHPR	African Commission on Human and Peoples' Rights
Additional Protocol to the Budapest Convention	Additional Protocol to the Convention on Cybercrime, Concerning the Criminalization of Acts of a Racist and Xenophobic Nature Committed through Computer Systems 2003
Belém do Pará Convention	Inter-American Convention on the Prevention, Punishment and Eradication of Violence against Women 'Convention Belém do Pará' 1994
Budapest Convention	Convention on Cybercrime of the Council of Europe 2001
CAT	United Nations Committee Against Torture
CEDAW	United Nations Convention on the Elimination of All Forms of Discrimination against Women 1979
CEDAW Committee	United Nations Committee on the Elimination of Discrimination against Women
CERD	United Nations Committee on the Elimination of Racial Discrimination
CESCR	United Nations Committee on Economic, Social and Cultural Rights
CJEU	Court of Justice of the European Union
CoE	Council of Europe
CRC	United Nations Convention on the Rights of the Child 1989
CRPD	United Nations Convention on the Rights of Persons with Disabilities 2006

DSA	Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC
ECHR	European Convention on the Protection of Human Rights and Fundamental Freedoms 1950
ECJ	European Court of Justice
e-Commerce Directive	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')
ECommHR	European Commission on Human Rights
ECtHR	European Court of Human Rights
EU	European Union
GDPR	The General Data Protection Regulation 2016
GREVIO	Group of Experts on Action against Violence against Women and Domestic Violence (CoE)
IACmHR	Inter-American Commission on Human Rights
IACtHR	Inter-American Court of Human Rights
ICC	International Criminal Court
ICCPR	International Covenant on Civil and Political Rights 1966
ICERD	International Convention on the Elimination of All Forms of Racial Discrimination 1965
ICESCR	International Covenant on Economic, Social and Cultural Rights 1966
ICL	International Criminal Law
ICT	Information and Communications Technology
ICTR	International Criminal Tribunal for Rwanda
ICTY	International Criminal Tribunal for the Former Yugoslavia
IHL	International Humanitarian Law
ILO	International Labour Organization
IP	Internet Protocol
IRL	In Real Life
ISP	Internet Service Provider
Istanbul Convention	Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence 2014
Lanzarote Convention	Council of Europe Convention on Protection of Children against Sexual Exploitation and Sexual Abuse 2007

Maputo Protocol	Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa 2003
MNE Declaration	Tripartite Declaration of Principles Concerning Multinational Enterprises and Social Policy
NGO	Non-Governmental Organization
OAS	Organization of American States
OECD	Organisation for Economic Co-operation and Development
PACE	Parliamentary Assembly of the Council of Europe
TFEU	Treaty on the Functioning of the European Union (consolidated version of 2016)
UDHR	Universal Declaration of Human Rights 1948
UN	United Nations
UNCAT	United Nations Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment 1984
UNESCO	United Nations Educational, Scientific and Cultural Organization
UN HRC	United Nations Human Rights Committee
UNODC	United Nations Office on Drugs and Crime
URL	Uniform Resource Locator
VLOP	Very Large Online Platform
VPN	Virtual Private Network
WHO	World Health Organization

Chapter 1
Introduction

1.1 Background

The Internet is increasingly indispensable to various human activities in the public and private spheres, such as education, entertainment, work, commerce, information-seeking, and political debate, evidenced by the fact that approximately five billion people were using the Internet in 2021.[1] Although originally designed as a scientific research tool and military communications network, the Internet is in its current phase largely utilised as a social web, characterised by user-generated content and social networking.[2] Users may create, upload and modify their own content and interact with other users through a variety of mediums. Simultaneously, both businesses and governments aim to reduce costs and increase accessibility to information by gradually shifting services to the Internet.[3] News, books and research are available online, with the Internet being the primary means of accessing information for young people.[4] In view of these features, the Internet is a particularly important platform for social groups with limited access to the public sphere, for example, due to discrimination and/or socio-economic impediments, as a means of accessing information, building knowledge, expressing ideas and mobilising social and political change.[5] As noted by the UN Special Rapporteur on the Freedom of

[1] International Telecommunication Union, statistics, <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> Accessed 9 February 2022.

[2] Powell and Henry (2017), p. 33.

[3] Meyer (1999), p. 307.

[4] CoE, 'Internet Governance – Council of Europe Strategy 2016–2019: Democracy, Human Rights and the Rule of Law in the Digital World' (Adopted at the 1252th Committee of Ministers' Deputies Meeting on 30 March 2016), para. 11.

[5] Laidlaw (2015), p. 25.

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7_1

Expression, the Internet provides novel modes of citizen participation and is an essential vehicle for democracy.[6]

Given the importance of this sphere, access to information and communication technologies (ICTs) is increasingly addressed through international human rights law, primarily viewed as being embedded in the freedom of expression, both to impart and receive information.[7] For example, the European Court of Human Rights (ECtHR) considers the Internet to be one of the primary means through which individuals are able to exercise their freedom of expression.[8] Internet access is also viewed as important to such objectives as the right to development, democracy and gender equality.[9] Such positive aspects have similarly been addressed within the framework of women's international human rights law. Technological advances may, for example, aid in creating networks, bridge gaps of physical distance and serve to empower women, especially in rural areas with limited access to state services such as health-care, legal assistance and victim support.[10] The Internet also stimulates employment, education and political participation. Additionally, ICTs provide platforms for a new sexual citizenship, allowing women to explore areas of their sexuality that may not be allowed or considered appropriate in their cultural setting.[11] In a global study from 2013, 85% of women consequently

[6]UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27, paras. 20, 22.

[7]UNHRC, 'Resolution on the Promotion, Protection and Enjoyment of Human Rights on the Internet' (29 June 2012) UN Doc A/HRC/20./L.13; UNHRC, Res 22/6 on Protecting Human Rights Defenders (12 April 2013) UN Doc A/HRC/22/L.13; CoE, 'Recommendation CM/Rec (2007)16 of the Committee of Ministers to Member States on Measures to Promote the Public Service Value of the Internet' (Adopted by the Committee of Ministers on 7 November 2007 at the 1010th meeting of the Ministers' Deputies); IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013) OEA/Ser. L./V/II. CIDH/RELE/INF.11/13, para. 15.

[8]*Ahmet Yildirim v Turkey* App no. 3111/10 (ECtHR, 18 December 2012), para. 54.

[9]ICESCR, 'General Comment 13: The Right to Education (Art. 13)' (8 December 1999) UN Doc E/C.12/1999/10, para 6; European Parliament Recommendation of 26 March 2009 to the Council on Strengthening Security and Fundamental Freedoms on the Internet (2008/2160(INI)); Beijing Declaration and Platform for Action, Report of the Fourth World Conference on Women, Beijing, 4–15 September 1995, UN Doc A/CONF.177/20 and UN Doc A/CONF.177/20/Add.1, para. 33; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 67; CEDAW, 'General Recommendation No. 33 on Women's Access to Justice' (23 July 2015) UN Doc CEDAW/C/GC/33, para. 17 (d); CEDAW, 'General Recommendation No. 34 on the Rights of Rural Women' (4 March 2016) UN Doc CEDAW/C/GC/34, para. 75; UNHRC, '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, para. 53.

[10]UNGA, 'Intensification of Efforts to Eliminate all Forms of Violence against Women and Girls: Report of the Secretary-General' (30 July 2020) UN Doc. A/75/274, detailing how the Internet provides support services, for example, for victims of abuse.

[11]Hearn (2006), p. 952.

held that the Internet provided them with more freedom.[12] The Internet has thus been hailed—also by many feminist legal scholars—as a liberating sphere for women where the significance of gender roles is diminished.[13]

Nevertheless, the Internet is in certain respects a male sphere in terms its development, architecture, user demographics and regulation. This has gendered consequences both in relation to access and online content. Although the expansion of the Internet into social networking has increased the number of female users—with women currently representing the majority of users of the main social media platforms—a quantitative gender gap remains.[14] While women constitute approximately half of the web population in certain Western states, women in most societies have less access to the Internet.[15] With a more limited presence in the production and use of ICTs, women accordingly have the least impact on their purpose and content. The gender gap on the Internet has in international human rights law mainly been approached from such a quantitative standpoint, with policies and programmes aiming to increase women's access to technology.[16] Meanwhile, the gendered "participation divide"—meaning the harmful nature and effect of online content and communication—has received less attention in this area of law.[17]

Empirical studies indicate that women, to a higher degree than men, are subjected to a variety of gender-based offences online, which may cause physical, psychological or financial harm.[18] Although ICT-related offences are commonly approached as a coherent category of contraventions of domestic and/or international law, they

[12] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call' (23 October 2015), <https://www.broadbandcommission.org/publication/cyber-violence-against-women/> Accessed 28 March 2022, p. 15.

[13] Haraway (1991); Wajcman (2004), p. 7.

[14] McGraw (1995), p. 495.

[15] UN Women, 'Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development' (2018), pp. 101–102; OECD, 'Bridging the Digital Gender Divide: Include, Upskill, Innovate' (2018), <http://www.oecd.org/internet/bridging-the-digital-gender-divide.pdf> Accessed 5 June 2020, p. 22; UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 2; International Telecommunication Union, gender statistics, <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> Accessed 25 May 2021.

[16] *See*, for example, Wajcman et al. for UN Women, 'The Digital Revolution: Implications for Gender Equality and Women's Rights 25 Years after Beijing' (August 2020), <https://www.unwomen.org/en/digital-library/publications/2020/08/discussion-paper-the-digital-revolution-implications-for-gender-equality-and-womens-rights> Accessed 31 March 2022; UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017) UN Doc. A/HRC/35/9; European Parliament resolution of 21 January 2021 on Closing the Digital Gender Gap: Women's Participation in the Digital Economy (2019/2168(INI)).

[17] Powell and Henry (2017), p. 253.

[18] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 2; European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey:

involve a wide array of violations differing as to the identity of the perpetrator, the technology platform used, the nature of the violence and the harm suffered. Online offences are generally conducted through speech, be it in words or images, which in certain instances result in physical acts of harm. Individual harmful acts include online harassment, such as receiving unsolicited nude images, defamation and the disclosure of personal information, images and videos.[19] Sexual violence may also be perpetrated through or be facilitated by ICTs. This includes being coerced into sending nude photographs or to engage in virtual sexual activities, such as undressing in live video chats, or physical acts offline.[20] Beyond individual instances of gender-based harm, certain forms of content portray and enforce stereotyped roles of women as sex objects and as the inferior sex. For example, online pornography is easily accessible, with videos containing demeaning and aggressive sexual behaviour prevalent on mainstream pornography sites.[21] Sexist hate speech—denigrating women as a group—is also increasingly widespread.[22] The UN Special Rapporteur on Violence against Women has as a result contended that the Internet has become 'a site of diverse forms of violence against women, in the form of pornography, sexist games and breaches of privacy'.[23] Although such gender-based offences pre-date the Internet, they are facilitated, exacerbated or expressed in new forms in this forum. This has broader implications for democracy, as it may cause female users to retreat from the Internet—in several respects analogous to a public space—and skewers information available online.

While *access* to the Internet is encompassed in certain pre-existing human rights, it is also clear that the full range of international human rights law applies equally online as offline. [24] Calls are also increasingly made for the alignment of technological development—including Internet architecture—with human rights law norms.[25] Nevertheless, rights equivalence does not entail a uniform application void of contextual considerations but rather a transposition that maintains the

Main Results' (2014); CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016).

[19] Halder and Karuppannan (2011), p. 387.

[20] Additionally, several reports note an increase in human trafficking, facilitated by ICTs, through using the web for recruitment, advertisement and sale of people. *See* Hughes (2014), pp. 1–8; CoE, 'Group of Specialists on the Impact of the use of New Information Technologies on Trafficking in Human Beings for the Purpose of Sexual Exploitation: Final Report' (16 September 2003) EG-S-NT (2002) 9 Rev.

[21] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 7; Vera-Gray et al. (2021).

[22] CoE, 'Seminar Combating Sexist Hate Speech: Report', 10–12 February, EYC, Strasbourg, p. 6.

[23] UNHRC, '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), para. 25.

[24] *See* Sects. 2.2.2.2 and 2.2.2.3.

[25] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021) UN Doc. A/76/258, para. 84;

efficacy of regulation. This must address how Internet architecture and social norms affect the prevalence and nature of gender-based offences and, in turn, the content of rights. Accordingly, while international human rights law places obligations on states to prevent a range of gender-based offences and gender stereotyping, the Internet as the area of abuse generates challenges in the recognition of harm as violations. These are both practical and ideological in nature. Internet architecture, including its end-to-end design, user anonymity, the principal role of Internet intermediaries and its transboundary nature, exacerbates gender-based harm and undermines effective regulation. Meanwhile, the structure of international human rights law limits areas subject to regulation and agents of liability. The theoretical foundation of rights—such as the freedom of expression and the right to privacy—as well as the concept of "harm", construct narrow boundaries that in certain instances exclude or underestimate the gravity of gendered violations and online harm, creating gaps in protection. The value of the Internet in turn affects the balancing in conflicts of rights and interests. That is, the online context affects the legal approach to what is considered harmful, the scope of protection, and the delineation and enforcement of liability, with gendered effects.

In this regard, an analysis of gender-based harm on the Internet from the viewpoint of international human rights law, in combination with feminist legal theories, will be useful for advancing the legal and theoretical frameworks on this issue. The book will first explore general challenges in the application of international human rights law to the Internet. Subsequently, a more detailed analysis follows of whether and how select forms of online conduct and materials—categorised as constituting or contributing to gender-based harm—are addressed within international human rights law. This will indicate potential gaps in protection, related to both the space of offences—the Internet—and the gendered nature of the harm.

1.2 Research Aims and Questions

The main aim of the book is to examine whether and how international human rights law applies to select gender-based offences online. In doing so, broader questions are addressed, such as the manner in which the online context affects the scope and content of rights and obligations. This considers such aspects as the assessment of harm, legal concepts, balancing in conflicts of interests and the regulation of intermediary liability. Potential gendered effects are considered. That is, are international human rights law provisions, including their theoretical underpinnings, inclusive of women's experiences when applied to the context of the Internet? Within this overall theme, the three main chapters will consider separate but interlinked research questions:

UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), paras. 42, 44.

1. *How does international human rights law apply to the Internet? Which contextual elements should be considered in relation to online gender-based offences?* This chapter develops the premise of the study and includes the following assertions: (1) international human rights law applies to the Internet. This encompasses state obligations to protect individuals against gender-based violations and to combat gender stereotyping; (2) the forms of online speech and acts analysed in the book constitute or contribute to gender-based violations; (3) online gender-based violations are both similar and distinct from non-virtual offences; and (4) the online/offline equivalence of human rights law requires a contextual and gender-sensitive approach, that is, consideration of how Internet architecture and online social norms exacerbate and facilitate gender-based offences, and affect the means of regulation.
2. *Are there general attributes embedded in international human rights law that impede the regulation of online gender-based harm?* The chapter analyses theoretical and structural constraints *vis-à-vis* both gendered harm and Internet regulation, which may explain potential gaps evinced in relation to question 3. The review considers the theoretical and legal approach to the concept of harm. The scope and values of select human rights are analysed, which informs the approach to harm, the content of obligations and the balancing of rights and interests, in addition to the online/offline coherence of rights. The section on liability addresses obligations and means of regulating online users—balanced against the right to anonymity—Internet intermediaries and media publishers in international human rights law. Suggestions on re-interpretations of concepts and theories to enhance the protection against gender-based online harm will be made where applicable.
3. *Do the reviewed forms of harmful online content and conduct contravene international human rights law and, if so, what is the scope of state obligations?* In cases where conduct at a general level is considered to be in violation of international human rights law, I will assess how the Internet as the platform for violations affects the interpretation of rights and the balancing in conflicts of rights in relation to such offences. Where the harmful conduct or speech is not clearly regulated in international human rights law, the analysis will explore whether the Internet affects the need for regulation. Examining gender-based violations in their online form may also reveal general gaps of protection in this body of law. Subsequent to an overview of state obligations *vis-à-vis* the selected forms of harm, I will consider whether the context of the Internet affects, or should affect, the content of obligations in view of the technological and social attributes of this forum. That is, both in terms of regulation and the content of obligations the applicability of human rights law provisions to the Internet will be evaluated.

1.3 Methods, Theories and Materials

A combination of a legal dogmatic method and feminist legal methodology is used in the book. The former is employed by structuring and applying relevant provisions in international human rights law, that is, a presentation of *de lege lata*, interspersed in all research questions. This involves the examination of a broad range of international law sources, such as select international human rights law treaties; case law; soft law documents and academic scholarship.

As Internet-based offences, in the main, are not explicitly prohibited in international treaties, primarily the applicability of general international human rights law provisions in select treaties will be assessed. The treaties applied in the book vary depending on the subject matter but mainly involve the European Convention on the Protection of Human Rights and Fundamental Freedoms (the ECHR), the American Convention on Human Rights (ACHR), the International Covenant on Civil and Political Rights (ICCPR), the United Nations Convention of the Elimination of All Forms of Discrimination (CEDAW) and regional treaties on violence against women and women's rights, given their relevance to the subject matter.[26] Other treaties are mentioned where pertinent, for example, the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), in relation to hate speech.[27] Limited case law exists in the area of Internet regulation and international human rights law, mainly of the ECtHR. Nevertheless, case law on human rights law, in general, by a broader set of actors is also included. Still, much focus lies on the European system for human rights and it should be noted that the overview consequently presents a limited perspective on the matter, particularly in relation to the contentious area of legitimate restrictions on the freedom of expression. Given the novelty of the subject matter and evident gaps in regulation, much soft law material is also presented, such as reports by UN special rapporteurs, Council of Europe (CoE) recommendations as well as general comments and concluding observations by UN treaty bodies. Such sources are used as indications of emerging approaches in this area of law as well as plausible influences in the interpretation of human rights law provisions.

Meanwhile, feminist theories review the modalities of power, which produce gendered social relations. For feminist *legal* theorists, the law as the source of power is examined. Accordingly, although law—including international law—is presented

[26] European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), 4 November 1950, 213 UNTS 222, entered into force 3 September 1953; American Convention on Human Rights (1969) O.A.S.T.S. 36, 1144 UNTS 123, entered into force 18 July 1978; Convention on the Elimination of All Forms of Discrimination against Women (1979), GA Res 34/180, 18 December 1979, 1249 UNTS 13, entered into force 3 September 1981; International Covenant on Civil and Political Rights (1966) GA Res 2200A(XXI), UN Doc. A/6316, 999 UNTS 171; entered into force 23 March 1976.

[27] International Convention on the Elimination of All Forms of Racial Discrimination (1965), 21 December 1965, GA Res 2106 (xx), Annex, 20 UN GAOR Supp (No 14) at 47, UN Doc. A/6014 (1966), 660 UNTS 195; entered into force 4 January 1969.

as neutral, it may serve the interests of the powerful, to silence or suppress other narratives and conceal or reproduce relations of gender domination.[28] The construction of the law may as such compound harmful behaviour through *de facto* impunity. More specifically, the feminist legal method of "asking the woman question" involves analysing whether '. . .the law fails to take into account the experiences and values of women' and 'how existing legal standards and concepts might disadvantage women'.[29] Such an approach has similarly been employed as an aspect of gender-mainstreaming in international human rights law.[30] This method is a means of analysing whether gaps and inconsistencies exist in the international human rights law regime from the perspective of gender equality, that is, whether the framework presents areas of non-existing or limited protection for women or whether applicable regulations have gendered effects. As the Internet is a relatively recent technological medium, a regulatory gap concerning this forum is evident. However, as international human rights law norms are contextually neutral, the analysis of potential gaps and the gendered effects of regulations is rather conducted from the viewpoint that the application of provisions must be functional and effective in the context of the Internet, from the standpoint of substantive gender equality.

In order to identify such gaps in the legal framework, a range of feminist legal theories are presented in various chapters of the book. As a first step, feminist theories provide analytical tools for determining what constitutes harm from a legal perspective and what forms of offences are gendered. This includes a discussion of both individual and group-based transgressions as instances of substantive gender inequality. From this standpoint, feminist legal theories assess the gendered construction and effect of public international law. In this regard, legal, philosophical and feminist theories on select rights are discussed, more specifically the freedom of expression and the right to privacy. The aim is to expose the ideological foundations and gendered interpretations of rights, with a view to applying rights in a gender-sensitive manner, while maintaining their underlying values. Such theories also inform the analysis of the causes and consequences of online offences and how the regime of international law responds, limits or exacerbates them, beyond the individual perpetration of a particular crime. Specific feminist theories on the gendered development, content and use of the Internet—cyberfeminism—informs the contextual approach to international human rights law.

In doing so, "gender" is a central concept in the assessment of the structural causes and consequences of violations. "Gender" involves the social and cultural construction of attributes associated with men and women, such as gender roles and

[28] Finley (1989), p. 888; Charlesworth and Chinkin (2000). As argued by Catharine MacKinnon, law reproduces the experiences of men, which in effect ensures control of women's bodies. *See* MacKinnon (1983), p. 644.

[29] Bartlett (1990), p. 837.

[30] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006) UN Doc. A/61/122/Add.1, para. 41, calling for the recognition of women's particular experiences.

stereotypes, in contrast to "sex" as a biological element.[31] Gendered offences involve victims of either sex, and "gender" is thus not synonymous with "women". In fact, by connecting "gender" to sex, it is implied that gender is a fixed characteristic.[32] At the same time, the prevalence and nature of harm is in certain regards distinct in relation to women as a group. For example, the forms of gendered violations discussed in the book are most commonly perpetrated against women and are often of a distinct nature *vis-à-vis* female victims, with violations frequently involving sexual objectification and harm to sexual autonomy.[33] Gender-neutrality may thus be inimical to assessing potential gaps and inconsistencies in current regulations. Accordingly, a tension can be noted at the international level between developing gender-neutral standards that nonetheless consider gender as a component, and rights focusing particularly on women.[34] Both approaches are valuable but may be more or less relevant depending on the issue and context.

Increasingly, "gender-based harm against women" is employed as a frame of reference. This addresses the gendered causes and consequences of violations, while also recognising the commonalities among victims.[35] This will also be the approach in this book. The focus on women allows for the application of the existing legal framework involving women's human rights, such as CEDAW and the regional human rights law treaties on the topic. Meanwhile, in considering the gender-based nature of violations, such offences are addressed as part of a continuum of violence against women.[36] However, this does not limit the scope to solely cisgender women, but also includes individuals who identify as such. Furthermore, as empirical studies indicate that gender intersects with age, ethnicity and sexual orientation, in addition to career paths (women in prominent positions) and ideologies (women expressing feminist ideas), in increasing the risk of being subjected to online harassment and violence, intersectionality will be mentioned where relevant.[37] Nevertheless, intersectional theories and methods are not comprehensively applied in the book.

---

[31] Office of the Special Advisor on Gender Issues and Advancement of Women UN, 'Gender Mainstreaming: Strategy for Promoting Gender Equality', (August 2001), <http://www.un.org/womenwatch/osagi/pdf/factsheet1.pdf> Accessed 11 February 2022; Otto (2013), p. 206.

[32] Charlesworth (2011), p. 31. By not addressing men and male gender identities, it is also implied that male characteristics are natural.

[33] *See* statistics in Sect. 2.2.2.4.4.2.

[34] For example, the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (the Istanbul Convention) (2011), CETS No. 210, entered into force 1 August 2014, which contains both gender-neutral (domestic violence) and gender-specific (violence against women) provisions.

[35] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19' (14 July 2017) UN Doc. CEDAW/C/GC/35; UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 11. For a discussion on this issue, *see* Boyle (2019), p. 20.

[36] Boyle (2019), p. 24.

[37] For the concept of intersectionality, *see* Crenshaw (1989). For studies on victims of online harm, *see* Sect. 2.2.2.4.4.2.

The theories employed, whether involving feminist or rights theories, have been selected on the basis of their relevance to the subject matter and consequently do not provide a complete representation of theoretical frameworks. Nevertheless, multiple theories on these issues will be presented in order to ensure a diversified analysis. Feminist theories are manifold and do not embody a cohesive approach to gender equality, public international law or the Internet. For example, the scope of the concept of "equality" and the degree of special accommodation for women divides feminist legal scholars.[38] This is also apparent in the context of the Internet. However, feminist theories maintain a similar core, also beyond the aim of gender equality. Feminist legal theories are mainly contextual, primarily through the use of "practical reasoning", that is, the interpretation of norms in relation to facts, experiences and contexts.[39] This is an aspect of both cultural[40] and radical feminism[41] and involves bringing to light underlying political and moral implications of rights and considering 'specific, real-life dilemmas posed by human conflict'.[42] The need for contextualisation in relation to actual social conditions rather than the application of abstract principles is deemed necessary for reasons of equality.[43] Accordingly, the values of rights and freedoms cannot be considered in isolation, for example, when interests collide. For instance, in cases of conflicts between the freedom of expression and the right to privacy, the former commonly prevails as it is viewed as a fundamental aspect of democracy, which may be disadvantageous to women. Although practical reasoning is mainly employed in individual cases, a contextual approach will be used at a more general level in the book, to consider the interpretation of rights and values in light of "gender" and the context of the Internet.

Sections of the book also include brief excursions into theories beyond the legal sphere, such as linguistic philosophy. This is included in chapters on the social effects of speech, connected to the legal concept of harm, as well as the categorisation of speech as specific forms of human rights law violations, for example, as discrimination. It is primarily relevant in relation to hate speech and certain forms of pornography, with regulation premised on the harmful effects on particular social groups. Arguments for considering such theories in the interpretation of rights are thus made.

It should be noted that certain harmful acts and forms of speech discussed in the book have been extensively analysed in previous academic scholarship, in relation to

[38] Whereas liberal feminists generally approach law as a neutral and genderless regime, cultural and radical feminist scholars emphasise the distinction of female characteristics and experiences in various areas of life, requiring accommodation through distinct regulation by the state. *See* Bowman and Schneider (1998), p. 251.

[39] Katharine Bartlett defines abstract reasoning as involving 'universal principles and generalizations', criticising it for glossing over 'specific, real-life dilemmas posed by human conflict'. *See* Bartlett (1990), pp. 836, 849. *See* also Boyle (1992); Baines (2009), p. 31.

[40] Bartlett (1994), p. 12.

[41] Grahn-Farley (2011), p. 112.

[42] Bartlett (1990), p. 850.

[43] Baines (2009), p. 31.

specific bodies of law. While new technologies have generated novel forms of gender-based offences, such as image-based sexual abuse, several of the examined violations existed prior to the Internet. For example, the elements of and obligations *vis-à-vis* sexual harassment are generally well-developed at the domestic level, primarily in the context of the workplace and in educational settings.[44] There is also a wealth of feminist scholarly work on the harmful effects of pornography.[45] However, these theories were primarily developed before the globalisation of Internet use. Similarly, the international human rights law approach to sexual violence is well documented, but again, to a limited extent in the context of the Internet.[46] Accordingly, the book makes a valuable contribution to the development of effective regulation in this area by addressing the issue comprehensively, combining the framework of international human rights law with theoretical perspectives, applied to the specific context of the Internet.

1.4 Delimitations

In pursuing the aims of the book, certain delimitations are necessary. Substantive gender equality on the Internet involves both equal access to this forum as well as protection against online gender-based violations. Although these aspects are interlinked, the book focuses on the latter. Furthermore, the book mainly concerns violations against adult victims. In comparison to offences against children, it is an emerging and theoretically more contested area of ICT-facilitated harm. As will be discussed, states espouse diverging ideological and theoretical approaches to restricting the freedom of expression and the minimum point of agreement at the international level concerning cybercrime is the prohibition on the production, distribution and downloading of child exploitation images and grooming, and certain forms of data interference.[47] Children are categorised as a vulnerable group in international human rights law, and states thus attract broader positive obligations to protect this group.[48] Nevertheless, child exploitation images and grooming will be discussed in certain chapters by way of analogy in relation to theoretical discussions on certain forms of online content, such as adult pornography.

[44] *See*, for example, MacKinnon (1979); MacKinnon and Siegel (2004). *See* more in Sect. 4.3.2.

[45] MacKinnon (1985); Dworkin (1989). *See* more in Sect. 4.5.

[46] *See*, for example, Sjöholm (2017), ch. 6; Edwards (2010); McGlynn and Munro (2011).

[47] The Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (Lanzarote Convention) (ETS No. 201) 25 October 2007; Convention on Cybercrime of the Council of Europe (Budapest Convention) (ETS No. 185) 23 November 2001.

[48] United Nations Convention on the Rights of the Child (1989) GA Res 44/25, annex, 44 UN GAOR Supp (No 49) at 167, UN Doc. A/44/49 (1989), 1577 UNTS 3; entered into force 2 September 1990; African Charter on the Rights and Welfare of the Child, OAU Doc. CAB/LEG/24.9/49 (1990), entered into force Nov. 29, 1999; *A v the United Kingdom* (1999) 27 EHRR 611, para. 22.

Although gender-based offences are perpetrated through a variety of ICTs, the book will solely address harm committed through the Internet. Furthermore, gender-based harm on the Internet involves a broad category of offences. Addressing the general attributes of such online offences is beneficial in highlighting similarities in their underlying causes and broader consequences. This accentuates the need for comprehensive state measures, such as eliminating harmful gender stereotyping. Nevertheless, the offences also differ in terms of their nature, harm, means of prevention and regulation and thus also require differentiation. Although sections of a general nature in the book are of relevance regardless of such factors, the scope has thus been narrowed to cover select online gender-based offences. This concerns various forms of sexual violence, including rape, forced nudity and the non-consensual distribution and receipt of intimate images. The book also analyses harassment, including sexual harassment, defamation and the disclosure of private information without consent. Furthermore, regulation of sexist hate speech and harmful pornography is examined. Accordingly, the monograph will not consider, for example, the crime of stalking or domestic violence committed through ICTs, nor the facilitation by ICTs of such gendered violations as enforced prostitution or human trafficking.[49] Nevertheless, certain violations addressed in the book may be components also of such offences.

Furthermore, online transgressions of domestic and international law are perpetrated by states, companies and private individuals. Offences on the Internet may accordingly engage multiple areas of law at both the domestic and international level, such as intellectual property law, EU law, domestic criminal law, and public international law. The same activity may be subject to concurrent rules in multiple legal regimes, engendering different types of obligations depending on the identity of the actor and the activity in question. Material considered unlawful mainly at the domestic level, for example, engages international private law and transborder criminal law *vis-à-vis* individuals, generating complex issues involving jurisdiction. Meanwhile, when offences are identified as violations of international human rights law, state obligations arise. However, certain chapters include comparative sections on other areas of law, for example, international criminal law (ICL) and EU law, and brief mentions are made of select domestic criminal law provisions for the purpose of exemplifying regulation. Furthermore, whereas the delineation of state jurisdiction *vis-à-vis* online offences is one the primary practical impediments to regulation and enforcement of domestic laws, such discussions will largely be excluded for reasons of space.

The means of addressing online gender-based offences through international human rights law are multiple. With respect to the prevention of such violations, states mainly incur positive obligations to protect individuals. Meanwhile, negative

[49] It has been argued that the concept of human trafficking should not be limited to that of the physical person but also include images, that is, virtual trafficking. *See* Council of Europe, 'Group of Specialists on the Impact of the use of New Information Technologies on Trafficking in Human Beings for the Purpose of Sexual Exploitation: Final Report' (2003), p. 50. However, such an approach will not be taken in the book.

obligations of non-interference, for example, in relation to surveillance and the collection of private information, limit the choice of means of protection. Obligations to protect may involve the adoption of effective criminal and/or civil legislation, investigating complaints, prosecuting offenders and undertaking operational measures. Means of preventing harm also extend more broadly, for example, to education in schools and informing the general public, police and judiciary on the harmful effects of gender stereotypes both In Real Life (IRL) and online.[50] Many international organisations and scholars thus place particular importance on state measures addressing the root causes of gender-based violence, regardless of the forum.[51] Nevertheless, the chapter on specific offences primarily focuses on legislative state obligations as the content of such is more developed in international human rights law. Other types of obligations will be discussed where relevant. Furthermore, although the transformation of pervasive cultural patterns of masculinity arguably requires mainly social rather than legal solutions,[52] social norms also acquire validity and direction from legislation. The law thus also holds an expressive function—shaping social norms by shifting the meaning of behaviour.[53]

References

Baines B (2009) Contextualism, feminism, and a Canadian woman judge. Fem Leg Stud 17:27–42
Bartlett K (1990) Feminist legal methods. Harv Law Rev 103:829–888
Bartlett K (1994) Gender law. Duke J Gend Law Policy 1:1–20
Bowman CG, Schneider EM (1998) Feminist legal theory, feminist lawmaking, and the legal profession. Fordham Law Rev 67:249–271
Boyle C (1992) The role of the judiciary in the work of Madame Justice Wilson. Dal Law J 15:241–260
Boyle K (2019) What's in a name? Theorising the inter-relationships of gender and violence. Fem Theory 20:19–36
Charlesworth H (2011) Talking to ourselves? Feminist scholarship in international law. In: Kouvo S, Pearson Z (eds) Feminist perspectives on contemporary international law: between resistance and compliance? Hart, Oxford, pp 17–32
Charlesworth H, Chinkin C (2000) The boundaries of international law: a feminist analysis. Manchester University Press, Manchester

[50] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 30; CEDAW, 'General Recommendation No. 36 on the Right of Girls and Women to Education' (16 November 2017) UN Doc. CEDAW/C/GC/36, para. 72; UNHRC, '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), para. 110; UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 28.

[51] *See* Nussbaum (2010), pp. 68, 85–87; UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 7.

[52] Nussbaum (2010), p. 85.

[53] Citron and Penney (2019), p. 2321.

Citron DK, Penney J (2019) When law frees us to speak. Fordham Law Rev 87:2317–2335
Crenshaw K (1989) Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chic Leg Forum 1989: 139–167
Dworkin A (1989) Pornography: men possessing women, 4th edn. Plume, New York
Edwards A (2010) Violence against women under international human rights law. Cambridge University Press, Cambridge
Finley L (1989) Breaking women's silence in law: the dilemma of the gendered nature of legal reasoning. Notre Dame Law Rev 64:886–910
Grahn-Farley M (2011) The politics of inevitability: an examination of Janet Halley's critique of the criminalisation of rape as torture. In: Kouvo S, Pearson Z (eds) Feminist perspective on contemporary international law, between resistance and compliance? Hart, Oxford, pp 109–132
Halder D, Karuppannan J (2011) Cyber gender harassment and secondary victimization: a comparative analysis of the United States, the UK, and India. Vict Offenders 6:386–398
Haraway D (1991) Simians, cyborgs and women: the reinvention of nature. Routledge, New York
Hearn J (2006) The implications of information and communication technologies for sexualities and sexualised violences: contradictions of sexual citizenship. Polit Geogr 25:944–963
Hughes D (2014) Trafficking in human beings in the European Union: gender, sexual exploitation, and digital communication technologies. Sage Open, pp 1–8
Laidlaw E (2015) Regulating speech in cyberspace: gatekeepers, human rights and corporate responsibility. Cambridge University Press, Cambridge
MacKinnon C (1979) Sexual harassment of working women: a case of sex discrimination. Yale University Press, New Haven
MacKinnon C (1983) Feminism, Marxism, method, and the state: toward feminist jurisprudence. Signs 8:635–658
MacKinnon C (1985) Pornography, civil rights and speech. Harv Civ Rights Civ Liberties Law Rev 20:1–70
MacKinnon C, Siegel R (eds) (2004) Directions in sexual harassment law. Yale University Press, New Haven
McGlynn C, Munro V (eds) (2011) Rethinking rape law: international and comparative perspectives. Routledge, London
McGraw D (1995) Sexual harassment in cyberspace: the problem of unwelcome E-mail. Rutgers Comput Technol Law J 21:491–518
Meyer C (1999) Women and the internet. Tex J Women Law 8:305–324
Nussbaum M (2010) Objectification and internet misogyny. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 68–90
Otto D (2013) International human rights law: towards rethinking sex/gender dualism. In: Davies M, Munro V (eds) The Ashgate research companion to feminist legal theory. Routledge, London, pp 197–215
Powell A, Henry N (2017) Sexual violence in a digital age. Palgrave Macmillan, London
Sjöholm M (2017) Gender-sensitive norm interpretation by regional human rights law systems. Brill, Leiden
Vera-Gray F et al (2021) Sexual violence as a sexual script in mainstream online pornography. Br J Criminol 61:1243–1260
Wajcman J (2004) Technofeminism. Polity Press, Oxford

Chapter 2
The Internet: A Gendered Space

2.1 Introduction

The aim of the following chapter is to establish a framework through which the main research questions will be explored. This broadly considers the relationship between the Internet and international human rights law, before addressing protection against gender-based offences in particular. Within this area of law, the Internet may be addressed from two perspectives: is there a human right to access the Internet and how do human rights apply online? The first issue affects the degree of protection required. If access to the Internet is considered a human right *per se* or an aspect of existing rights, obligations arise in ensuring substantive equality in terms of content and communication. It also indicates the value of the Internet which, for example, has an impact on balancing exercises in conflicts of rights. Meanwhile, a contextual approach to the application of international human rights law standards takes into consideration how rights translate to the online context. Regulation relevant to preventing gender-based offences is in this regard presented.

The final section addresses the Internet as a gendered sphere. The premise is that, although the causes of online offences against women are similar to gender-based violations in general, the characteristics of the Internet influence the pervasiveness, the effect and the regulation of harmful conduct. This, in turn, must affect the application of human rights online, including the scope of rights, the resolving of conflicts of rights and the content of state obligations.

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7_2

2.2 A Rights-Based Approach to the Internet

2.2.1 The Role for International Human Rights Law in Internet Regulation

The value of applying international human rights law to the Internet and the question of *how* this legal regime may intervene is informed by the features, use and aim of this sphere. As will be viewed, the complexity in regulating the Internet relates to a variety of factors, addressed throughout the book. The following section will thus provide a brief overview of Internet governance.

The Internet is a global system that interconnects computer networks and other electronic devices for the transmission of information.[1] Developed in the early 1970s by academic researchers, it was subsequently adopted as a military device by the US government, with the aim of creating a communications network that could transfer large data files between the government and government-sponsored research laboratories.[2] Although initially a state sponsored project, it is increasingly privatised. The main impetus for commercialisation was the introduction of the personal computer and the creation of the Ethernet in the late 1980s, which linked personal computers through local area networks. With the launch of the World Wide Web in 1991, transmission of and access to information over the Internet was globalised, beyond the earlier private networks of the 1980s.

With the development of Web 2.0, both users and intermediaries play a significant role in controlling online content. The Internet remains decentralised and is largely operated by private sector entities, such as Internet Service Providers (ISPs), website operators, and social media platforms. There is both public (state) and private (company) control, but in practice regulation often occurs through website terms of use, ISP contracts, moderator filtering, browser and server code, and social norms of behaviour.[3] Terms of service are provided by virtually all ISPs, social network providers, and blogging services.[4] At the same time, users maintain substantial control over their communication and dissemination of information. Depending on the website, users can control the audience of their published content, for instance, through privacy settings on social media. Such interaction occurs through a wide range of media, including Usenet newsgroups, the World Wide Web, email, webcams, live video conferencing, streaming video, peer-to-peer servers, and file sharing programs. Communication is in certain instances one-to-one, that is, between private individuals and in other channels one-to-many or many-to-many, involving distribution to multiple individuals. Simultaneously, government services and information are increasingly transferred online, entailing a

[1] Radu (2019), p. 44.

[2] ibid., p. 46.

[3] Netanel (2000), p. 400.

[4] Benedek and Kettemann (2013), p. 103.

deprivatisation of the Internet and an increase of its public service value. In view of its regulation, control, purpose and operation, the Internet thus has attributes of both a public and private sphere. It has accordingly been categorised as a private sphere with a public dimension or, reversely, a '. . .progressive public sphere made up of private spheres where. . .issues of public interest are discussed'.[5] This affects where and how states may intervene. For example, the public sphere connotations of the Internet have been noted by UN bodies, the CoE and the EU as a cause for increasing regulation and control, from the viewpoint of gender equality.[6] In contrast, state interference is subject to certain restraints in private pockets.

The a-territorial nature of the Internet—unbound and unregulated by states—influenced the early discourse on its use as a means of fostering a global community and a democratic space.[7] The view of the Internet as a separate sphere from the physical, and even the textual—a cyberspace—was emphasised by cyberlibertarian scholars such as John Barlow.[8] The lack of regulation ensured the possibility of genuine democracy, where private citizens may 'speak truth to power', creating a marketplace of ideas and direct communication between individuals.[9] It was construed as a digital public sphere, allowing for a broadening of contributors to the public discourse, with limited consideration of its private space implications and the real-life effects of virtual communication.

Like cyberlibertarians, early feminist scholars approached the Internet with anticipation. Scholarly literature was often blended with science fiction, envisioning the Internet as a liberating space and a means of reducing social constructs of gender and sexuality.[10] It was presumed that the digital revolution would cause a decline of traditional institutional sources of power, such as the state and patriarchal beliefs.[11] Particularly user anonymity was viewed as a means of allowing individual expressions of multiple—often unexplored—sides of the self, creating a space for questioning gender relations.[12] The Internet accordingly held the potential of being a genderless sphere free of patriarchy, and a space for constructing identities outside the regular realms of social control, such as family, work and school.[13] Arguably, '[w]hen gender boundaries defining masculinity and femininity begin to loosen,

[5] ibid., p. 103.

[6] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021) UN Doc. A/76/258, para. 91; European Parliament resolution of 11 September 2018 on measures to prevent and combat mobbing and sexual harassment at workplace, in public spaces, and political life in the EU (2018/2055(INI)), para. C; PACE Resolution 2177 (2017) 'Putting an End to Sexual Violence and Harassment of Women in Public Space'.

[7] Schulte (2011), p. 731.

[8] Barlow (2019).

[9] Benedek and Kettemann (2013), p. 103.

[10] Haraway (1991). *See* also Schulte (2011), p. 732 for an overview.

[11] Wajcman (2004), p. 7.

[12] Daniels (2009), p. 121.

[13] Turkle (1995), ch. 8.

even in virtual space. . .they unsettle patriarchal heterosexual traditions'.[14] Female empowerment would not be limited to the online space but also strengthen gender equality offline.[15]

Given these distinctive characteristics of the Internet, the anticipation was that a *sui generis* regulatory framework would govern this area.[16] According to libertarian scholars, traditional legal concepts do not apply to cyberspace.[17] The arguments are in part ideological—with cyberspace viewed as a new type of direct democracy, outside the reach of territorial governments—and in part based on the challenges associated with technological infrastructure. Cyberspace does not have its own territory, people or a right to self-determination and thus no independent rules stemming from internal or external sovereignty.[18] Meanwhile, fundamental aspects of sovereignty linked to territoriality—being power, legitimacy, effects and notice—do not exist in cyberspace.[19] Nor are there convincing claims to designate cyberspace a *res communis*,[20] despite assertions that the Internet is a global public good and that sovereignty claims projected onto this sphere may destroy or compromise its values.[21] Furthermore, social norms are, arguably, different on the Internet, leading to an inapplicability of the rule of law and a lack of acceptance of domestic laws by users.[22] The normative base would instead be a social contract created and implemented by individuals in the form of self-regulation.[23]

However, given the extensive role and impact of the Internet in society, states maintain a significant interest in its governance, in practice exercising domestic jurisdiction over online activities in accordance with traditional principles of jurisdiction in public international law.[24] At the same time, given its architecture and principal features, state regulation cannot function in its regular form. First of all, in the legal regime developing at the international level, Internet governance is often addressed as a shared responsibility, in view of the range of actors with varying degrees of *de facto* control. Secondly, the system of national laws and a transnational Internet is, to a degree, irreconcilable. As data flows across international borders, is stored simultaneously in multiple jurisdictions and may have different legal status depending on the state, the issue of jurisdiction is highly complex. The options for states are thus to make domestic law more transnational (through the harmonisation of regulation) or online activity less transnational (through blocking, filtering,

[14] Eisenstein (1998), p. 90.
[15] Haraway (1991); Daniels (2009), p. 109.
[16] Tsagourias (2015), p. 13.
[17] Barlow (2019).
[18] Tsagourias (2015), p. 24. *See* also Miller (2003), p. 234.
[19] Post and Johnson (1996), p. 1370; Kohl (2007), p. 157.
[20] Tsagourias (2015), p. 28; Tully (2014), p. 190.
[21] Kohl and Fox (2017), p. 23.
[22] Lessig (2006), p. 124.
[23] Fidler (2015), p. 97.
[24] Kittichaisaree (2017), p. 23.

monitoring or use of geolocations), that is, transferring territorial boundaries into cyberspace.[25]

Although a process is underway within the EU and the CoE to harmonise the approach to the illegality of online content, international regulation specific to this issue is limited.[26] The tendency is thus currently to divide the Internet into distinct spheres governed by separate regulations rather than an open network, diminishing core cyberlibertarian ideals.[27] Certain central aspects of the Internet, such as its end-to-end design and global nature, are thus being compromised.[28] From the perspective of international human rights law, this development is primarily considered to undermine the protection of the freedom of expression. Technological fragmentation reflects the offline world, with varying approaches to the content of rights embodied in domestic laws. However, as publishers increasingly need to code each item of information published on the Internet and set filters in accordance with national standards, intermediaries often opt to restrict speech in order to avoid the burden of identifying the location of end-users.[29] The territorialisation of the Internet may also result in a spillover of national laws into other states, which forces conformity to the most restrictive common denominator, constraining the freedom of expression.[30] At the same time, the lack of international harmonisation of Internet regulation impedes efforts to restrict gender-based harm. With already limited domestic regulation of such offences, it weakens prescriptive incentives and enforcement of domestic laws.[31] Internet intermediaries may also circumvent strict domestic regulations limiting certain types of speech by placing servers in host states with less restrictive laws, and users can evade borders through Virtual Private Networks (VPN) or the darknet.

International human rights law thus fills an important role is this regard. This area of law comprises a universal categorisation of conduct and content as harmful—indicating who and how it harms—on the basis of core values. It also provides

[25] Kohl (2007), p. 28; Schultz (2008), p. 804.

[26] For example, the Convention on Cybercrime of the Council of Europe (the Budapest Convention) (ETS No. 185) 23 November 2001; the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (Lanzarote Convention) (ETS No. 201) 25 October 2007; 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; Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020); Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (2011) OJ L335/1.

[27] Kohl and Fox (2017), p. 5; Gagliardone et al. for UNESCO (2015), p. 56.

[28] Berger (2017), pp. 28, 38; Schultz (2008), p. 805; Crockett (2016), p. 167.

[29] Nyberg (2004), p. 684; Benedek and Kettemann (2013), p. 92.

[30] Smith (2017), p. 136.

[31] Reed (2012), p. 36; Kohl (2007), p. 259.

elements relevant for balancing competing interests, including hierarchies of values in conflicts of rights. Finally, it directs states in terms of obligations, including regulation of non-state actors, be it individuals or Internet intermediaries. In fact, it is increasingly argued that Internet architecture and technological development should be aligned with human rights law values, in view of the substantial effect of information and communication technologies (ICTs) on individuals. Nevertheless, as will be viewed, aspects of the utopian ideology remain, in the form of self-regulation by intermediaries, and a certain *laissez faire* approach in international law towards Internet regulation. The preservation of Internet architecture and its effective operation is a significant factor in rights interpretation, with more limited consideration of harmful online content and communication. The usefulness of the regime of international human rights law in regulating gender-based online offences—balanced against the respect for the freedom of expression—thus requires a context- and gender-sensitive approach, with the principle of equality in focus.

2.2.2 A Gender Equal Internet

2.2.2.1 The Principle of Equality

Equality is one of the main justifications for the development of the regime of international human rights law and remains an overarching principle in this area of law.[32] The principle stems from values such as social justice, freedom and dignity and is a fundamental element of democracy.[33] However, the role of the principle in this body of law varies. It is employed in multiple ways, including (1) as a preambular objective in international human rights law treaties;[34] (2) as a descriptive function of the scope and application of human rights[35] and (3) codified in

[32] For example, the IACtHR has stated that 'the notion of equality springs directly from the oneness of the human family and is linked to the essential dignity of the individual'. *See Proposed Amendments to the Naturalization Provisions of the Constitution of Costa Rica*, Advisory Opinion OC-4/84, IACtHR Series A No 4 (19 January 1984), para. 55. According to the Court, the principles of equality and non-discrimination constitute *jus cogens*, since both the national and international public order rests on such pillars. *See Juridical Condition and Rights of the Undocumented Migrants*, Advisory Opinion OC-18/03, IACtHR Series A No 18 (17 September 2003), para 101. The African Commission has stated that non-discrimination and equality are 'essential to the spirit of the African Charter'. *See Purohit and Moore v the Gambia*, ACmHPR, Communication No. 241/01 (29 May 2003), para. 49.

[33] Clifford (2013), p. 444.

[34] It is a general feature of preambles in human rights law treaties, that is, a principal purpose of the human rights law regime, alongside dignity.

[35] Ensuring that everyone has an equal right to human rights provisions, that is, it plays a procedural function in how human rights must be applied.

substantive provisions in human rights law treaties.[36] Through the latter two functions, it is also employed as an interpretative guide on the content of rights.[37] There is thus no autonomous right to equality *per se*, but rather provisions on non-discrimination, equality before the law or equality *vis-à-vis* specific human rights. However, it can be argued that in the application of the principle of equality by international and regional courts and treaty bodies, a substantive right to equality has developed in practice.[38]

Employing the concept of equality as a prism to assess aspects of gender and the Internet includes considering equal access *to* the Internet and equality *in* this forum. In relation to the latter, it establishes a framework for placing obligations on states to prevent and remedy gender-based violence and gender stereotyping. Furthermore, theories on gender equality inform the assessment of whether and how specific online activities are harmful and gender-based. Additionally, the concept of equality can be used to guide the interpretation of rights and conflicts of norms, primarily involving the freedom of expression and the right to privacy, in order to ensure a gender-sensitive approach to the Internet. The impact of the values associated with particular rights may be similarly addressed from an equality perspective. Gender equality is also employed in the feminist legal method, as the basis for evaluating potential gaps and inconsistencies in legal frameworks, from the viewpoint of disadvantageous effects on women as a group. Additionally, increasing calls are made to ensure the principle of equality in technological development and Internet architecture, including taking into account how products and services affect different social groups.[39] Accordingly, the concept of gender equality will be addressed in various chapters of the book.

Nevertheless, complications arise in both defining equality and concretising measures to achieve it, both in theory and law. Although there is a broad consensus in feminist theories for the elimination of inequality, a core *essentialist/constructivist* division is evident.[40] This affects the approach to female agency and autonomy and informs the categorisation of which types of content and conduct are

[36] Clifford (2013), p. 431. For example, Art. 26 of the International Covenant on Civil and Political Rights (1966) GA Res 2200A(XXI), UN Doc. A/6316, 999 UNTS 171; entered into force 23 March 1976 (ICCPR), on the right to equality before the law; and Art. 3 of International Covenant on Economic, Social and Cultural Rights, GA Res. 2200A (XXI), 21 UN GAOR Supp (No 16) at 49, UN Doc. A/6316 (1966), 993 UNTS 3, entered into force Jan. 3, 1976 (ICESCR) and the ICCPR, respectively, on the equal rights of women and men.

[37] Clifford (2013), p. 444.

[38] ibid., p. 442.

[39] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017) UN Doc. A/HRC/35/9, para. 13.

[40] It should be noted that these feminist theories share an assumption of an objective truth on which the law can be based. In contrast, postmodern feminist criticism is part of the critical legal studies movement that considers law to be indeterminate and non-objective. Postmodern feminists reject identity politics overall and assume that both "sex" and "gender" are constructed categories. *See* Levit and Verchick (2016), p. 10.

considered harmful. Essentialists view sex and gender as biological constructs and thus innate human characteristics. Meanwhile, a constructivist approach holds that gender and, according to some, also sex, stem from social factors, with the possibility of changing socially sanctioned cultural perceptions on gender that inevitably lead to oppression.[41] Cultural feminists ascribe to the first category while, generally, liberal and radical feminists adhere to the latter. *Which* gender differences should be recognised and accommodated varies depending on the feminist theoretical approach. It may, for example, involve biological, social or cultural characteristics.

In relation to the concept of equality, a liberal theoretical approach to women's human rights ascribes to the concept of formal equality and dictates that law should treat people, who are similarly situated, alike.[42] Liberal feminist theory thus emphasises biological similarities between the sexes, on the basis of liberal ideals of equal citizenship and opportunity.[43] Meanwhile, a substantive approach recognises that men and women are different in certain areas and that gender-neutral laws may produce unequal effects. This view encompasses the application of indirect discrimination and focuses on the equality of results rather than solely on form.[44] It thus adheres to the concept of *de facto* equality and requires special measures or differences in treatment, with a transformative aim.[45] Consequently, equality is not to be treated *without* regard to sex but requires freedom from systematic subordination *because* of sex.

For example, *cultural* feminists approach equality from a substantive standpoint, on the premise that men and women are different, with regard to both sex and gender, and that female characteristics are undervalued.[46] It thus finds its basis in essentialism, although a constructivist strand also exists.[47] For instance, Carol Gilligan, a psychologist and leading scholar in the development of cultural feminist theory, held that women reason with an ethic of care in interactions with other people, whereas men reason with an ethic of rights, emphasising the autonomy of the person.[48] The individualistic view of the right-holder, with rights emanating from the ethical value of autonomy predominant in human rights law, is thus rejected,

[41] Palazzini (2012), p. 16.

[42] Bowman and Schneider (1998), p. 251; Littleton (1991), p. 35.

[43] Williams (1991), p. 26; Levit and Verchick (2016), p. 12. However, certain differences are, to a degree, acknowledged, such as reproductive functions. *See* Charlesworth (1996), p. 559.

[44] *See*, for example, CEDAW, 'General Recommendation No. 25 on Article 4, paragraph 1, of the Convention on the Elimination of All Forms of Discrimination against Women, on Temporary Special Measures' (2004) UN Doc. A/59/38, para. 8.

[45] Charlesworth and Chinkin (1993), p. 148.

[46] Fineman (2005), p. 17; Littleton (1991), p. 36.

[47] Halley (2006), p. 59.

[48] Gilligan (1982).

since it is not representative of women's experiences, whose lives are rather associated with "connectedness".[49]

Meanwhile, certain feminists reject the traditional model of non-discrimination and suggest a "disadvantage model", where the effects of a law or policy are not analysed in relation to the individual, but rather at a group level.[50] A substantive equality approach and disadvantage model is evident in, for example, *dominance* theory (or radical feminism), which rejects the sameness-difference debate on the basis that both the above-mentioned theories maintain a male reference point. Rather than presuming that men and women are biologically similar or different, dominance theorists analyse the difference in *power* between men and women, without allowing men to be the norm against which the particular conduct is measured.[51] The focus of the theory is on male exploitation and the expropriation of women's sexuality as the main tools of subordination, evident through the disproportionate number of female victims of such abuses.[52] The main question for radical feminists is consequently whether laws, policies or practices perpetuate, facilitate or reinforce women's subordination.[53]

Although international human rights law in certain instances concretises the concept of gender equality, through the categorisation of violations as gender-based and in the content of state obligations, the approach is often *ad hoc*-based. From this patchwork, different strands of the principle appear, aligned with certain feminist legal theories. At a general level, it can be noted that equality and non-discrimination are considered distinct but intertwined concepts in international human rights law.[54] Whereas non-discrimination entails the prohibition on arbitrary difference in treatment—such as distinction, exclusion, restriction or preference—equality may require broader strategies, such as affirmative action measures, to achieve its goal.[55] Traditionally, the preferred approach to equality in the international human rights law regime, both explicitly and in practice, is that embodied by liberalism.[56] That is, the focus lies on ensuring formal equality through prohibiting direct discrimination, which entails that individuals in similar positions must be treated equally.[57]

[49] Radacic (2008a), p. 266.

[50] Mahoney (1994), p. 442; Radacic (2008b), p. 850.

[51] MacKinnon (1987), pp. 37–40.

[52] ibid.

[53] Radacic (2008a), p. 267.

[54] Arnardóttir (2002), p. 7.

[55] According to the IACtHR: 'States are obliged to take affirmative action to reverse or change discriminatory situations that exist in their societies to the detriment of a specific group of persons'. *See Juridical Condition and Rights of the Undocumented Migrants*, Advisory Opinion OC-18/03, IACtHR, para. 104.

[56] Nussbaum (1997), p. 2.

[57] Meyerson (2007), p. 175.

Nevertheless, obligations to ensure substantive equality are, to an increasing extent, developing in this area of law. Most relevantly, the United Nations Convention of the Elimination of All Forms of Discrimination (CEDAW) explicitly requires states to undertake measures to achieve substantive gender equality. This includes ensuring that '. . .all human beings, regardless of sex, are free to develop their personal abilities, pursue their professional careers and make choices without the limitations set by stereotypes, rigid gender roles and prejudices'.[58] Article 2 places broad obligations on states to ensure both *de jure* and *de facto* equality.[59] Meanwhile, regional human rights law courts have only to a limited degree applied this approach, although it is progressively employed. In the main, the Inter-American human rights institutions have affirmed an obligation of ensuring *de facto* equality, particularly *vis-à-vis* vulnerable groups.[60] In the 2019 case of *Volodina v Russia*, involving domestic violence, the European Court of Human Rights (ECtHR) also held that '[s]ubstantive gender equality can only be achieved with a gender-sensitive interpretation and application of the Convention provisions that takes into account the factual inequalities between women and men and the way they impact women's lives'.[61] Similarly, gender-based violence and gender stereotyping have been categorised at the international level as constituting gender discrimination and transgressions of the equality principle *per se*.[62] Nevertheless, regional human rights law courts, especially the ECtHR, in the main evaluate cases of alleged discrimination in isolation from social and cultural contexts.[63] Although offences against women are progressively recognised as violations of international human rights law, such as the right to privacy or the prohibition on inhuman or degrading treatment, harm to women's dignity and autonomy are thus still often viewed as disconnected events,

[58] CEDAW, 'General Recommendation No. 28 on the Core Obligations of States Parties under Article 2 of the Convention on the Elimination of All Forms of Discrimination against Women' (16 December 2010) UN Doc. CEDAW/C/2010/47/GC.2, para. 22.

[59] ibid., paras. 16, 24. More specifically, affirmative action is an obligation in Art. 4 of CEDAW, requiring the adoption of temporary special measures to accelerate *de facto* equality.

[60] *See Case of Artavia Murillo* et al. *v Costa Rica* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 257 (28 November 2012). The ECtHR allows, but does not oblige, states to adopt such measures. *See*, for example, *Stec and Others v the United Kingdom* (2006) 43 EHRR 47.

[61] *Volodina v Russia* App no 31261/17 (ECtHR, 9 July 2019), para. 111. Similarly, the IACtHR has held that each situation merits a different state response and different treatment in light of the American Convention, including affirmative action obligations. *See Proposed Amendments to the Naturalization Provisions of the Constitution of Costa Rica*, Advisory Opinion OC-4/84, IACtHR, para. 56. *See,* also, *Case of the Girls Yean and Bosico v Dominican Republic* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 156 (8 September 2005), para. 141; *Case of Artavia Murillo* et al. *v Costa Rica* (Preliminary objections, merits, reparations, and costs), IACtHR, para. 293.

[62] *See* Sect. 2.2.2.4.

[63] There are, however, exceptions. For example, *D. H. and Others v the Czech Republic* (2008) 47 EHRR 3. Calls are thus frequently made for regional human rights law courts to recognise also the structural aspects of violations against women. *See,* for example, Timmer (2011); Radacic (2008b).

taking place in the private sphere of relationship conflicts and beyond the realm of state policies.[64] This is also evident in relation to online offences.

Given the fluctuating approach to equality by human rights law courts and treaty bodies, the feminist theories mentioned enrich the analysis of international human rights law *vis-à-vis* gendered harm in general, and in the application to the Internet in particular. The various approaches to the concept of equality are not necessarily in conflict in terms of Internet regulation but may be applied to different aspects involving gender, be it access to the Internet, the involvement of women in technological development and education, and protection against online harm. Both *de jure* and *de facto* equality are relevant in this regard. Nevertheless, the focus on effectively combating gender-based violations and stereotyping requires a pronounced substantive equality approach, in view of the differences in online user experiences of men and women and the nature of the harm. Such an approach is also noticeable in recent soft law documents on this issue by UN and CoE bodies.[65] From a theoretical perspective, such difference may relate to either biological or social factors. For example, the cultural feminist approach considers difference in how men and women use and communicate on the Internet, whereas dominance feminism is a relevant foundation for the analysis of gender-based violence, hate speech and pornography. The latter theory not only provides arguments for categorising such instances of harm as inequality *per se* or violations of specific human rights law norms, the focus on disadvantage also calls for an analysis of whether international law perpetuates, facilitates or reinforces female subordination. Through such theoretical perspectives, tools for establishing a more gender equal space may appear.

2.2.2.2 A Human Right to Access the Internet?

Despite the significant social impact of the Internet and its public sphere attributes, due to its relative novelty, no explicit right to access this medium is included in major international human rights law treaties, apart from the International Convention on the Protection and Promotion of the Rights and Dignity of Persons with Disabilities (CRPD), which obliges states to 'promote the availability and use of new technologies, including information and communications technologies. . .suitable

[64] WHO, 'Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-Partner Sexual Violence' (2013), p. 31.

[65] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 8; CoE, Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', adopted on 20 October 2021; UNHRC, '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.

for persons with disabilities. . .'.[66] Given the multifunctional use of the Internet, the UN Human Rights Council declared Internet freedom a basic human right in 2013, in particular as a means of ensuring the freedom of expression, thus encouraging states to foster access to the Internet.[67] A range of UN bodies and regional organisations have also announced obligations for states to take progressive steps to promote universal access, with special consideration of vulnerable groups, and the development of technological infrastructure.[68] While these sources employ terminology implying positive obligations to access the Internet, it is phrased in language similar to obligations *vis-à-vis* socio-economic rights, such as "facilitating", "subsidising", "taking steps to ensure" and "promoting availability".

In other instances, it is emphasised that the Internet is merely a means of ensuring existing rights, that is, of instrumental value. Accordingly, '[t]he Internet is a tool for imparting and receiving information. Internet access has no purpose independent of this context'.[69] As a technological tool, it is subject to change and, similar to electricity, it is a utility rather than a right.[70] Several international organisations thus consider Internet access as an instrument in ensuring human rights in general and the freedom of expression in particular.[71] For example, the ECtHR has noted that the Internet has become '. . .one of the principal means by which individuals

[66] Art. 4 (1) (g) of the International Convention on the Protection and Promotion of the Rights and Dignity of Persons with Disabilities, GA Res 61/106, Annex I, UN GAOR, 61st Sess., Supp No 49, at 65, UN Doc. A/61/49 (2006), entered into force 3 May 2008.

[67] Accordingly, '[a]ccess to and use of information technologies and the media of one's choice, including. . .the Internet, should be promoted and facilitated at the national level, as an integral part of the enjoyment of the fundamental rights to freedom of opinion and expression.' *See* UNHRC, Res 22/6 on Protecting Human Rights Defenders (12 April 2013) UN Doc A/HRC/22/L.13, para. 7. Though this was interpreted by some as an announcement by the UN of a new right, it should rather be viewed as a recognition of the importance in ensuring Internet access as policy. *See* also UNHRC, 'Resolution on the Promotion, Protection and Enjoyment of Human Rights on the Internet' (29 June 2012) UN Doc A/HRC/20./L.13.

[68] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013) OEA/Ser.L./V/II. CIDH/RELE/INF.11/13, para. 15; UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)' (2011) UN Doc. CCPR/C/GC/34, para 15; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (20 April 2010) UN Doc A/HRC/14/23, para. 37; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27, para. 85.

[69] Fidler (2015), p. 110.

[70] According to this view, there is a risk of human rights law inflation. *See*, for example, De Hert and Kloza (2012). Arguably, there is also no real need to introduce an independent right to Internet communication as the existing human right to freedom of expression in international human rights law treaties already encompasses this medium. *See* Benedek and Kettemann (2013), p. 32.

[71] *See* CoE, 'Recommendation CM/Rec(2007)16 of the Committee of Ministers to Member States on Measures to Promote the Public Service Value of the Internet' (Adopted by the Committee of Ministers on 7 November 2007 at the 1010th meeting of the Ministers' Deputies)*;* UNHRC, 'Report of the Working Group on the Issue of Discrimination against Women in Law and in Practice' (19 April 2013) UN Doc. A/HRC/23/50, para. 15. *See* also UNCHR, 'Report of the Special

exercise their right to freedom of expression and information, providing as it does essential tools for participation in activities and discussions concerning political issues and issues of general interest'.[72] As such, the ECtHR has affirmed that, under certain circumstances, a right to access the Internet exists for particular groups and that restricting means of communication interferes with the right to receive and impart information.[73] The blocking of entire sites, such as search engines, is an example of a breach of the freedom of expression.[74]

The freedom of expression is in turn considered a fundamental aspect of an open and democratic society, both in relation to imparting and accessing information. As such, the link to democracy is often highlighted in connection with a right to Internet access. Political participation is enhanced through discussion boards and blogs.[75] It allows people to impart and exchange ideas, knowledge and opinions, store vast amounts of information and has the potential to enhance tolerance between individuals from diverse cultures, backgrounds and beliefs. This includes vulnerable and marginalised groups and people in geographically remote or underdeveloped areas, and the Internet is thus important for pluralism in the public sphere.[76] Its integral role in calls for democracy was evident, for example, during the Arab Spring in 2011, where demonstrations were organised through social media.[77] Meanwhile, non-democratic regimes in certain instances suppress the free flow of information through ISP control.[78]

Internet access is also important for the fulfilment of economic, social and cultural rights. For example, the UN Special Rapporteur on the Freedom of Expression argues that 'the Internet boosts economic, social and political development, and contributes to the progress of humankind as a whole',[79] especially facilitating

Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015) UN Doc.A/HRC/29/32, para. 11.

[72] *Ahmet Yildirim v Turkey* App no. 3111/10 (ECtHR, 18 December 2012), para. 54.

[73] *Kalda v Estonia* App no 17429/10 (ECtHR, 19 January 2016). If the state is willing to allow prisoners access, it has to give a reason for refusing access to specific sites.

[74] *Ahmet Yildirim v Turkey* (ECtHR).

[75] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 13.

[76] CoE, 'Internet Governance – Council of Europe Strategy 2016–2019: Democracy, Human Rights and the Rule of Law in the Digital World' (Adopted at the 1252th Committee of Ministers' Deputies Meeting on 30 March 2016), para. 6; CoE, 'Recommendation CM/Rec(2011)8 of the Committee of Ministers to Member States on the Protection and Promotion of the Universality, Integrity and Openness of the Internet' (Adopted by the Committee of Ministers on 21 September 2011 at the 1121st meeting of the Ministers' Deputies), para. 2.

[77] Puddephatt for UNESCO, 'Freedom of Expression and the Internet' (2016), p. 19.

[78] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 30.

[79] ibid. para. 67. *See* also UNHRC, 'Resolution on the Promotion, Protection and Enjoyment of Human Rights on the Internet' (29 June 2012) UN Doc A/HRC/20./L.13, para. 2.

economic development for marginalised groups.[80] The Organization of American States (OAS) similarly considers that ICTs are essential for political, economic, social and cultural development, job creation, environmental protection and the reduction of poverty.[81] The Internet likewise enhances the right to education[82] and the right to participate in cultural life.[83] It is thus viewed as a facilitator of a range of human rights.

It should be noted that state obligations arise regardless of whether Internet access is considered an independent right or an aspect of pre-existing human rights. However, the content of obligations and the value attached to the Internet may be enhanced if construed as an autonomous right. Although beneficial from a gender perspective, an inflation of the value of the Internet could simultaneously generate an overcautious approach to its regulation. For example, it could tip the balance in favour of non-regulation in conflicts with other rights, such as in cases of harmful speech, further explored in Sect. 3.3.4. At the same time, with a rights-based approach to Internet access—irrespective of its status—principles of openness and equality follow.

As a consequence, obligations for states to ensure Internet access with a focus on women is developing. Access to ICTs is currently gendered, with fewer women using the Internet and contributing to its architecture and content.[84] Women are therefore more likely to be denied information, education and business opportunities provided in this sphere. Empirical studies indicate that the digital gender gap mainly stems from women's lack of education—such as illiteracy and inadequate language and computer skills—as well as economic constraints and a lack of access to equipment.[85] Nevertheless, women in the same age group and with similar levels

[80] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 62.

[81] OAS, 'Use of Telecommunications/Information and Communication Technologies to Build and Inclusive Information Society,' General Assembly of the Organization of American States, Asunción (Paraguay) (4 June 2014), AG/RES. 2859 (XLIV-O/14), p. 218; OAS, 'Advancing Hemispheric Initiatives on Integral Development', General Assembly of the Organization of American States, Santo Domingo (Dominican Republic) (14 June 2016) AG/RES. 2881 (XLVI-O/16), p. 71.

[82] ICESCR, 'General Comment 13: The Right to Education (Art. 13)' (8 December 1999) UN Doc. E/C.12/1999/10, para 6; European Parliament Recommendation of 26 March 2009 to the Council on Strengthening Security and Fundamental Freedoms on the Internet (2008/2160(INI)), para. O.

[83] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 76. *See* also Art. 15 of the ICESCR, providing a right to enjoy the benefits of scientific progress. *See* also Laidlaw (2015), p. 18.

[84] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 41; UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 4.

[85] Gurumurthy (2004), p. 23.

of education and income as men are still less likely to use the Internet.[86] As noted by the UN High Commissioner for Human Rights, the gender digital divide is not only a cause of human rights violations, but also a consequence.[87] Gender stereotypes, both in the public and private spheres, reinforce the digital divide.[88] The fear of harassment likewise affects the willingness to engage online.[89] Gender-based violence thus impedes the freedom of expression of women, also in relation to the Internet.[90] Internet access is consequently subject to issues of wealth and power, with social inequalities linked to income, education, age, gender, literacy and geography reproducing inequalities in cyberspace.[91] Accordingly, '[w]hen viewed in social terms, the virtual and real worlds are mutually constitutive: discrepancies in access to the internet both mirror and constitute inequalities in the world outside cyberspace'.[92] Meanwhile, the lack of access and limited online skills also increase women's vulnerability in preventing and responding to cyber attacks.[93]

In view of this imbalance, access to ICTs was recognised as a critical means in enhancing women's empowerment and gender equality as early as at the Fourth World Conference on Women in Beijing in 1995.[94] International and regional institutions have followed suit. For example, the UN Special Rapporteur on the Freedom of Expression and the UN High Commissioner for Human Rights affirm the need for states to empower women by enhancing skills and access to information technology.[95] Accordingly, '[g]iven that the internet has become an indispensable tool for realising a range of human rights, combating inequality, and accelerating

[86] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 6.

[87] ibid., para. 17.

[88] OECD, 'Bridging the Digital Gender Divide: Include, Upskill, Innovate' (2018), <http://www.oecd.org/internet/bridging-the-digital-gender-divide.pdf> Accessed 5 June 2020, p. 23.

[89] Gurumurthy (2004), p. 23; OECD, 'Bridging the Digital Gender Divide: Include, Upskill, Innovate' (2018), p. 22.

[90] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 12.

[91] Warf (2017), p. 155; Kohl and Fox (2017), p. 21; Tully (2014), p. 188.

[92] Warf (2017), p. 163.

[93] Bardall (2017), p. 105.

[94] Beijing Declaration and Platform for Action, Report of the Fourth World Conference on Women, Beijing 4–15 September 1995, UN Doc A/CONF.177/20 and UN Doc A/CONF.177/20/Add.1, Platform for Action, Strategic Objective B.3.

[95] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (20 April 2010), para. 47; UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 13. Similarly, the Special Rapporteur on the Freedom of Expression of the IACmHR encourages states to promote access that is universal, equitable and affordable throughout states' territory, in addition to training and education and '. . .places a priority on ensuring equitable access when it comes to gender. . .'. *See* IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 16.

development and human progress, ensuring universal access to the Internet should be a priority for all states'.[96] The United Nations Committee on the Elimination of Discrimination against Women (the CEDAW Committee) has also in several general recommendations noted the importance of access to new technology for women's empowerment[97] and the UN Special Rapporteur on Violence against Women considers equal access an aspect of women's freedom of expression, the right to participate in political decision-making and non-discrimination.[98]

However, there is a shift in the discourse on women's access to ICTs. Whereas the previous approach mainly construed women as recipients of information, a more recent focus is on women as *contributors* and controllers of content.[99] This enhances the broader aims of informational pluralism and democratic participation. The gender-based digital divide hampering access to ICTs for women affects the substance of information available. Women are less visible in the media in general, be it as subjects of stories and reports, as reporters or participants, and women's perspectives are thus less likely to be broadcasted.[100] Additionally, according to cultural feminist theories, women contribute to social environments in different ways than men.[101] This is affirmed by sociological studies, indicating that men tend to dominate web-based discussion boards, both in terms of initiating topics of discussion as well as using rhetorical intimidation to control the conversation.[102] With more limited participation, the digital realm becomes more masculine. In a space where mainly men contribute and shape the content of information, gender stereotypes are not only replicated but also exacerbated.[103] Accordingly, women's involvement in developing Internet architecture and producing online content strengthens the prevention of online gender-based offences.

---

[96] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 85.

[97] CEDAW, 'General Recommendation No. 33 on Women's Access to Justice' (23 July 2015) UN Doc CEDAW/C/GC/33, para. 17 (d); CEDAW, 'General Recommendation No. 34 on the Rights of Rural Women' (4 March 2016) UN Doc CEDAW/C/GC/34, para. 75.

[98] UNHRC, '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), para. 53. The 2030 Agenda for Sustainable Development similarly recognises that the development of ICTs has the potential to accelerate human progress and encourages states to increase access to promote women's empowerment (targets 5 b and 9 c). *See* UN Women, 'Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development' (2018), p. 101; UN HRC, 'Resolution on the Promotion, Protection and Enjoyment of Human Rights on the Internet' (18 July 2016) UN Doc. A/HRC/RES/32/13.

[99] Gurumurthy (2004), p. 11.

[100] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 43.

[101] West (1988), p. 15.

[102] Herring (1996b), p. 486; Herring et al. (1995), p. 68.

[103] UNHRC, '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), para. 25.

As a consequence, the CEDAW Committee encourages the increased participation of women both as consumers and entrepreneurs in relation to ICTs.[104] The Beijing Declaration holds that '[u]ntil women participate equally in both the technical and decision-making areas of communications and the mass media. . .they will continue to be misrepresented and awareness of the reality of women's lives will continue to be lacking'.[105] This is linked to the benefits of the Internet to democracy, with the Inter-American Commission on Human Rights (IACmHR) noting that '[m]aximizing the number and diversity of voices that are able to participate in the public debate is both a means and an end of the democratic process'.[106] Furthermore, included in the freedom of expression is a right to receive a wide range of information and ideas.[107] Media plurality and qualitative aspects of user participation have thus been affirmed as positive obligations for states.[108] Although the focus on the Internet as a democratic space often highlights access to information, a similar emphasis must thus be placed on equal opportunities to participate and contribute content. Additionally, addressing a right to Internet access from a broad perspective not only generates obligations to ensure the availability and affordability of technological infrastructure and resources, but also protection from interpersonal harm denoted as human rights law violations.

[104] CEDAW, 'Concluding Observations on Austria' (1 May 2000) UN Doc. A/55/38 Part II, para. 223.

[105] The Beijing Declaration and Platform for Action (Platform for Action), para. 118.

[106] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 18.

[107] *See* also CERD, 'Concluding Observations on Croatia' (17 August 1998) UN Doc. A/53/18, para. 325, concerning ethnic groups. A CoE Recommendation from 2007 on measures to promote the public service value of the Internet similarly places obligations on states to ensure both Internet access and that ICT content is reflective of all regions, countries and communities in order to represent all peoples, nations, cultures and languages. *See* CoE, 'Recommendation CM/Rec(2007) 16 of the Committee of Ministers to Member States on Measures to Promote the Public Service Value of the Internet'.

[108] *See* IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 16; 'Joint Declaration on Freedom of Expression and the Internet', signed by the UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and Special Rapporteur on Freedom of Expression and Access to Information and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information on 1 June 2011, principle 5 a. *See* also *Kimel v Argentina* (merits, reparations and costs) IACtHR Series C No 177 (2 May 2008), para. 57; *Demuth v Switzerland* App no 38743/97 (ECtHR, 5 November 2002); CoE 'Recommendation CM/Rec(2016)1 of the Committee of Ministers to Member States on Protecting and Promoting the Right to Freedom of Expression and the Right to Private Life with Regard to Network Neutrality' (Adopted by the Committee of Ministers on 13 January 2016, at the 1244th meeting of the Ministers' Deputies).

2.2.2.3 Human Rights Online

Whereas the previous section affirmed that *access* to the Internet can be viewed as an aspect of a range of human rights, this part considers to what extent human rights law provisions apply *on* the Internet. Given the particular characteristics of the Internet, the adoption of specialised international treaties on the subject or a comprehensive approach to adapting existing rights to this sphere would be reasonable venues. However, different approaches in states on acceptable interferences with the freedom of expression—arising from factors such as political ideology, culture and morality—impede efforts of international cooperation.[109]

Concretely, there is limited agreement as to what types of online content and conduct are sufficiently harmful to warrant regulation. The sole treaties on the matter are the Convention on Cybercrime of the Council of Europe 2001 (the Budapest Convention) and the Protocol to the Convention on Cybercrime, Concerning the Criminalization of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (the Additional Protocol to the Budapest Convention), the former prohibiting activities such as hacking and the distribution of child-abuse images using a computer system, and the latter prohibiting hate speech, xenophobia and racism.[110] Special protection of children from being exposed to or exploited in injurious material online as well as regulation of terrorist-related activities on the Internet are also provided in a variety of EU directives.[111] Meanwhile, gender-based violence on the Internet was first addressed within the UN in 2006 by the Secretary-General in his in-depth study on violence against women, requesting more information on the use of ICTs and emerging forms of violations.[112] The topic has since been considered by several UN bodies[113] and regional human rights law organisations and

[109] Laidlaw (2015), p. 241.

[110] The Budapest Convention; the Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (ETS No. 189) 28 January 2003. Several states desired the inclusion of other content-related offences in the main convention, such as racist propaganda. However, as there was no consensus among states, it resulted in the adoption of an additional protocol. *See* CoE, 'Explanatory Report on the Convention on Cybercrime' (ETS No. 185) 23 November 2001, para. 35. There is also a proposal to develop a cybercrime convention through the UN. *See* UNGA, Resolution 72/247 on countering the use of information and communications technologies for criminal purposes, UN Doc. A/RES/74/247 (20 January 2020).

[111] Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography; Audiovisual Media Services Directive (2018); Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA (2017) OJ L88/6.

[112] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006) UN Doc. A/61/122/Add.1, para. 371.

[113] UNHRC, '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); UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 12.

courts, although still to a limited degree.[114] The recognition of the international human rights law implications of such violence is thus a relatively recent phenomenon and primarily developed in soft law documents. Accordingly, since explicit codification is limited, mainly general human rights law provisions are applied to the Internet.

In relation to *which* rights apply, it is continuously emphasised by international organisations such as the UN and the CoE that the full range of human rights applies equally online as offline.[115] Thus, although the phrasing of certain human rights law provisions affirms a direct applicability to the Internet, such as the freedom of expression in the International Covenant on Civil and Political Rights (ICCPR)—which applies regardless of frontiers and through any media of one's choice—this concerns human rights provisions in general.[116] In practice, existing provisions have been interpreted by various international bodies in a manner that incorporates expressions and acts on the Internet. For example, the ECtHR, the CEDAW Committee, the UN Human Rights Committee (UN HRC) and the UN Special Rapporteur on the Freedom of Expression have all affirmed that human rights law norms are fully applicable to technology-mediated environments, such as the Internet.[117]

As to *how* rights apply, there is less guidance, developed in an *ad hoc* manner in case law of regional human rights law courts and in soft law documents by UN treaty bodies. International human rights law provisions are purposefully abstract with rights given concrete content through international adjudication and state implementation. The evolutive treaty interpretation method favoured by regional human rights law courts/commissions and UN treaty bodies allows for a dynamic approach to the content of rights, bearing in mind the particular character of human rights law treaties in offering protection to individuals.[118] This special object and purpose

[114] For example, CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women'; *Buturuga v Romania* App no 56867/15 (ECtHR, 11 February 2020); European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', COM(2022) 105 final (8 March 2022).

[115] For example, various UN bodies have asserted that 'the same rights that people have offline must also be protected online'. *See* UN HRC, 'The promotion, protection and enjoyment of human rights on the Internet' (27 June 2016) UN Doc. A/HRC/32/L.20, para. 1; UNGA, Resolution adopted by the General Assembly on 18 December 2013: The right to privacy in the digital age (21 January 2014) UN Doc. A/RES/68/167, para. 3. *See* also CoE, 'Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to human rights for Internet users' (Adopted by the Committee of Ministers on 16 April 2014 at the 1197th meeting of the Ministers' Deputies).

[116] Art. 19 of the ICCPR.

[117] *Delfi v Estonia* (2014) 58 EHRR 29; *K.U. v Finland* (2009) 48 EHRR 52; *Buturuga v Romania* (ECtHR); UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para 12; UNCHR, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Abid Hussain' (30 January 2002) UN Doc. E/CN 4/2002/75, p. 6; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19' (14 July 2017) UN Doc. CEDAW/C/GC/35, para. 20.

[118] *Soering v the United Kingdom* (1989) 11 EHRR 439, para. 87; *The Effect of Reservations on the Entry Into Force of the American Convention on Human Rights (Arts. 74 and 75)*, Advisory

entails that progressive interpretations for the person's benefit may be necessary, in light of evolving ethical standards as well as social, technological and medical advances, which may have an impact on the substance of rights.[119]

Even if it is generally accepted that international human rights law extends to cyber activities, the content of rights and obligations must in certain instances be adapted in accordance with defining characteristics of this forum.[120] As such, '[i]n theory, human rights function independent of any technology. In practice, technologies affect whether and how individuals enjoy human rights'.[121] Thus, although it is considered preferable that there is broad uniformity between online and offline laws, equivalence does not entail that laws should be technologically neutral, but rather that they enforce similar values and principles.[122] Functional equivalence, rather than formal transposition, is appropriate.[123] Such an approach encourages a contextual approach in the interpretation of human rights law provisions, with a view to ensuring effective individual protection while maintaining the aim and values of rights. Regardless of the significance of the Internet, international law *per se* is thus generally technology neutral, not conferring on the Internet more privileges than enjoyed by other media, all the while being sensitive to the format.[124] *Which* features of Internet architecture and online culture affect the interpretation of rights and *how* will be explored throughout the book.

2.2.2.4 Protection Against Gender-Based Harm

2.2.2.4.1 Introduction

Given that international human rights law applies equally to the Internet, obligations to ensure substantive gender equality—including protection against gender-based violations and the elimination of gender stereotypes—also extend to this sphere. As

Opinion OC-2/82, IACtHR Series A No 2 (24 September 1982), para. 29; *Case of Artavia Murillo* et al. *v Costa Rica* (IACtHR), para. 173; *Centre for Minority Rights Development (Kenya) and Minority Rights Group International on behalf of Endorois Welfare Council v Kenya*, ACmHPR, Communication No. 276/2003 (25 November 2009), paras. 144–162; *Roger Judge v Canada*, Communication No. 829/1998, UN HRC, UN Doc. CCPR/C/78/D/829/1998 (5 August 2002), para. 10.3.

[119] *Christine Goodwin v the United Kingdom* (2002) 35 EHRR 18.

[120] UNODC, 'Comprehensive Study on Cybercrime' (February 2013), p. 109; Benedek and Kettemann (2013), p. 19.

[121] Fidler (2015), p. 94. Also the UN Special Rapporteur on the Freedom of Expression et al. argue that regulations concerning other forms of media, e.g. telephony or broadcasting, cannot be transposed to the Internet but special regulation is necessary. *See* 'Joint Declaration on Freedom of Expression and the Internet' (2011), principle 1 (c).

[122] Reed (2012), p. 38.

[123] ibid., p. 107.

[124] Barendt (2016), p. 139.

noted by the UN Special Rapporteur on Violence against Women, both individual offences on the Internet, such as harassment, and stereotyped portrayals of women must be addressed.[125] As opposed to approaching gender equality and the Internet from the standpoint of access, this considers equality in terms of online content. It includes gender-based harm through *communication*, that is, where individual women are subjected to gender-based violence through speech or acts, as well as harm in terms of *representation*, through the dissemination of content maintaining or exacerbating harmful gender stereotypes.[126] The following section will give an overview of the criteria for the categorisation of gender-based violations and gender stereotyping, and state obligations in this regard, while relating such to online offences. This is approached at a general level—as a framework—with more specific analyses of particular online offences provided in Chap. 4.

2.2.2.4.2 Gender-Based Violations

In international human rights law, certain types of violations are categorised as gender-based, with due regard of their causes, nature and effect. This approach acknowledges that offences have social as well as individual dimensions, that is, that persons suffer harm not only as individuals but also because they belong to a particular group, which may shape the nature of offences.[127] Considering violations such as sexual violence or harassment as merely instances of interpersonal violence rather than manifestations of gender subordination conceals the social arrangements causal to such problems.[128] It in effect individualises group harm.[129] The categorisation thus links such violations to the concept of equality and the prohibition on non-discrimination. Addressing harmful online behaviour through the existing framework on gender-based violence in international human rights law thus ensures that such acts are not approached as isolated, disconnected incidents, while drawing attention to the vulnerability of women to such forms of harm.

Furthermore, the categorisation may affect the assessment of harm, in that certain offences produce gender-based consequences. The effects of gender-based violence are in certain instances not recognised within the legal concept of "harm" in

[125] UNHRC, '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), para. 27.

[126] A similar delineation on harm exists in relation to the abuse of children online. *See* UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. 13.

[127] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 9.

[128] Catharine MacKinnon argues that sexual harassment cannot be adequately addressed through domestic provisions on tort, which concern individual wrongs, and thus do not address this social dimension. *See* MacKinnon (2007), p. 227.

[129] Conaghan (1996), p. 430.

domestic law nor considered sufficiently harmful to be encompassed by international human rights law provisions. Moreover, to ensure effective prevention of gender-based violations, the construction of state obligations must consider the structural causes of such violence as well as both individual and social effects. This may require specific types of obligations, including measures to eradicate gender stereotypes, for example, through public education, the training of police officers and judges, as well as the adoption of gender-sensitive legislation.

While certain international bodies have listed violations considered to be gender-based in specific contexts,[130] such compilations are not all-encompassing, with theoretical means of assessing the gendered aspects of offences more useful. Such a flexible approach is necessary in consideration of the fact that novel forms of gender-based harm arise, for example, through the development of new technologies.[131] Also, the gendered forms and effects of violations deemed gender-neutral must be recognised.[132] Thus, while certain types of violence, such as sexual violence are considered gender-based *per se*, other offences may be expressed in a gendered manner.

The concept of "gender-based violence" is gender neutral, that is, it relates to both sexes, which is also the case for gender-based offences beyond violence.[133] However, gender-based violations in general, as well as specific forms, primarily affect women, which is also recognised at the international level.[134] According to General

[130] UNCHR, 'Report of the Special Rapporteur on violence against women, its causes and consequences, Ms. Radhika Coomaraswamy, submitted in accordance with Commission on Human Rights Resolution 2001/49, Cultural Practices in the Family that are Violent towards Women', UN Doc. E/CN.4/2002/83 (31 January 2002): UNHRC, '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); ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', Adopted at the 21st Extra-Ordinary Session of the African Commission on Human and Peoples' Rights, held from 23 February to 4 March 2017 in Banjul, The Gambia, para. 58.

[131] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006), para. 105.

[132] For example, in relation to crime of torture. *See* UN HRC, 'Report of the Special Rapporteur on Torture and Cruel, Inhuman or Degrading Treatment or Punishment, Manfred Nowak', UN Doc. A/HRC/7/3 (15 January 2008).

[133] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 9; ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', para. 59: 'Any person regardless of their gender may be a victim of sexual and gender-based violence.'

[134] For example, in the Inter-American Convention on the Prevention, Punishment and Eradication of Violence Against Women (Belém do Pará Convention) (1994), 33 i.l.m. 1534 (1994); Istanbul Convention Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (the Istanbul Convention) (2011), CETS No. 210, entered into force 1 August 2014; Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (Maputo Protocol) (2000), Adopted by the 2nd Ordinary Session of the Assembly

Recommendation No. 19 of the CEDAW Committee, violence against women consists of offences directed at women *because* of their sex/gender or those that *affect* women disproportionately.[135] Versions of this definition have subsequently been included in regional human rights law treaties on gender-based violence.[136] Gender-based violence thus includes violations exclusive to women, on the basis of biological differences, for example, in relation to their reproductive capacity, or where gender is otherwise the catalyst. This frequently involves harm to the sexual autonomy of the individual and objectification of women. As noted above, feminist theories on equality provide multiple frameworks for assessing which violations are gendered in this regard.[137] Meanwhile, statistics demonstrate the disproportionate number of female victims in relation to various offences. It should be noted that offences other than violence may also be gender-based and the same typology as in General Recommendation No. 19 can be applied to infer a gender aspect also in relation to those.

The fact that the victim is a woman is not an indication in itself that a violation is gender-based as the offence may be unrelated to sex or gender. For example, the Inter-American Court of Human Rights (IACtHR) in the case of *Rios* et al. *v Venezuela*, which concerned limitations on the freedom of expression, including violence, applied the same standards of evaluation as in CEDAW General Recommendation No. 19.[138] The claimants held that the violent attacks on certain female journalists were gender-based since they were affected by the acts of violence in a different manner and to a greater proportion than male victims.[139] The Court noted that although female journalists were attacked, so were their male colleagues. The representatives had thus not proved how the attacks were 'especially direct [ed] against women' nor why women became a greater target of attacks 'based on their condition [of being women]'.[140] That is, both quantitative and qualitative gender differences are relevant.

---

of the Union, CAB/LEG/66.6; entered into force 25 November 2005 and CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women'.

[135] CEDAW, 'General Recommendation No. 19: Violence Against Women' (1992) UN Doc. A/47/38, para. 6; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 1. Meanwhile, the UN Special Rapporteur on Torture considers that gender-specific violence aims to "correct" behaviour perceived as non-consonant with gender roles and stereotypes or to assert or perpetuate male domination over women. *See* UNHRC, 'Report of the Special Rapporteur on Torture and other Cruel, Inhuman or Degrading Treatment or Punishment, Manfred Nowak' (15 January 2008) UN Doc. A/HRC/7/3, fn 7.

[136] Art. 3 (a) of the Istanbul Convention; Art. 1 of the Belém do Pará Convention; Art. 1 (j) of the Maputo Protocol.

[137] *See* Sect. 2.2.2.

[138] *Case of Rios* et al. *v Venezuela* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 194 (28 January 2009). *See* also *Case of Perozo* et al. *v Venezuela* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 195 (28 January 2009).

[139] *Case of Rios* et al. *v Venezuela* (preliminary objections, merits, reparations, and costs), para. 278.

[140] ibid., para. 279; *Case of Perozo* et al. *v Venezuela* (IACtHR), paras. 295–296. The same review was applied in the *Castro-Castro* case of the IACtHR, involving both men and women who were

While the category of gender-based human rights law violations encompasses a variety of transgressions with gendered causes and consequences, particularly robust state obligations have been affirmed in relation to gender-based *violence*. It is considered to impair or nullify women's enjoyment of human rights by contributing to their subordination and their lower level of political participation, education, skills and work opportunities, thereby exposing them to other risks such as the propagation of pornography and other forms of commercial exploitation.[141] A range of treaties, case law and soft law documents by regional human rights law courts and UN treaty-based bodies thus consider violence against women a form of sex discrimination. Explicit recognition is included in regional human rights law treaties related to violence against women.[142] Although not explicitly mentioned in CEDAW, such violence has been recognised as a form of discrimination by way of interpretation, in General Recommendations No. 19 and No. 35 of the CEDAW Committee.[143] In the latter recommendation, the Committee held that *opinio iuris* and state practice indicate that the prohibition on gender-based violence against women has evolved into a principle of customary international law.[144] Violence against women has also been considered an aspect of gender inequality in case law by regional human rights law institutions, primarily by the IACmHR/IACtHR and, to a more limited extent, by the ECtHR and the ACmHPR.[145] This mainly involves cases on sexual violence and domestic violence.

Gender-based violence is defined in General Recommendation No. 19 as encompassing '. . .physical, mental or sexual harm or suffering, threats of such acts, coercion and other deprivations of liberty'.[146] Regional human rights law treaties on violence against women identify such violence in a similar manner.[147] It should be noted that gender-based violence is not limited to physical harm but defined broadly to include, for instance, psychological abuse. The CoE Convention on Preventing and Combating Violence against Women and Domestic Violence (the

---

subjected to forced nudity, but women were deemed to be affected to a greater extent. *See Case of Miguel Castro-Castro Prison v Peru* (merits, reparations and costs) IACtHR Series C No. 160 (25 November 2006), paras. 303–308.

[141] CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 11.

[142] The Belém do Pará Convention; The Istanbul Convention; the Maputo Protocol.

[143] CEDAW, 'General Recommendation No. 19: Violence Against Women'; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', preceded by UNGA, 'Declaration on the Elimination of Violence against Women' (23 February 1994) UN Doc. A/RES/48/104.

[144] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 2.

[145] *See*, for example, *Opuz v Turkey* (2010) 50 EHRR 28; *Miguel Castro-Castro Prison v Peru* (merits, reparations and costs) IACtHR; *Case Egyptian Initiative for Personal Rights and Interights v Egypt*, ACmHPR, Communication No. 323/06 (1 March 2011).

[146] CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 6.

[147] Art. 2 of the Belém do Pará Convention, Art. 1 (j) of the Maputo Protocol; Art. 3 (a) of the Istanbul Convention. *See* also the UN Declaration on the Elimination of Violence against Women, Art. 2.

Istanbul Convention), for example, explicitly obliges states to prevent psychological violence, by criminalising threats or coercion.[148] As the definitions of gender-based violence are broad and apply regardless of context, online harm against women is encompassed. In CEDAW General Recommendation No. 35, it is in fact noted that gender-based violence '...occurs in all spaces and spheres of human interaction...and their redefinition through technology-mediated environments, such as contemporary forms of violence occurring in the Internet and digital spaces'.[149] Similarly, the Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO) of the CoE has affirmed the application of the Istanbul Convention to new technologies.[150] This entails that, depending on the form and circumstances, both speech-based and physical acts of violence on the Internet are included in the concept of gender-based violence.

As noted, violence against women is considered a *form* of discrimination while discrimination is the fundamental *cause* of such violence.[151] Addressing gender-based violence as a matter of discrimination thus requires comprehensive measures beyond remedying individual instances of violence. Given the connection to inequality, according to the CEDAW Committee, obligations to address violence against women are the same as in relation to discrimination in general, that is, to ensure that laws, policies, programmes, and procedures do not discriminate,[152] in addition to training state officials on gender-based violence, to investigate offences, to prosecute and punish offenders and provide reparations.[153] Nonetheless, gender-based violence is not only addressed as a matter of discrimination or inequality at the international level but has been included by way of interpretation in multiple other human rights, depending on the offence and level of harm, including the right to privacy and the prohibition on torture, inhuman or degrading treatment.[154] Similar positive obligations have been affirmed by regional human rights law courts and UN

[148] Art. 33 of the Istanbul Convention.

[149] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 20.

[150] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 18.

[151] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006), para. 31.

[152] Art. 2 (c) and (g) of the Convention on the Elimination of All Forms of Discrimination against Women (1979), GA Res 34/180, 18 December 1979, 1249 UNTS 13, entered into force 3 September 1981.

[153] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 23.

[154] *M. C. v Bulgaria* (2005) 40 EHRR 20 (rape); *Opuz v Turkey* (ECtHR) (domestic violence); *González* et al. *(Cotton Field) v Mexico* (preliminary objections, merits, reparations and costs) IACtHR Series C No. 205 (16 November 2009) (sexual violence/killing); *Equality Now and Ethiopian Women Laywers Association (EWLA) v Federal Republic of Ethiopia,* ACmHPR, Communication No. 341/2007, Adopted by the African Commission on Human and Peoples' Rights during the 19th Extra-Ordinary Session, from 16 to 25 February 2016, Banjul, The Gambia (sexual violence).

treaty bodies in relation to such rights. Depending on the form of violence, more specific obligations have also been formulated, for example, the definition of rape required in domestic criminal law.[155] The connection between discrimination and state negligence *vis-à-vis* such positive obligations have also been made.[156] This, for example, entails that gaps in domestic criminal or civil law in relation to online harm mainly affecting women, or existing laws that have detrimental gendered effects, may violate such obligations. In addition, a failure to effectively investigate online offences, for example, on the basis of the perceived immorality of the victim or a devaluing of the harm, may contravene human rights law standards. Nevertheless, as noted above, the specific content of positive obligations requires a contextual approach, in view of Internet architecture. This is explored in more detail in relation to specific offences in Chap. 4.

2.2.2.4.3 Gender Stereotypes

Gender-based human rights law violations are closely linked to gender stereotypes on female sexuality, reproduction and motherhood. While gender stereotypes produce gender-based violations, the pervasiveness and impunity *vis-à-vis* such offences further entrench stereotypes.[157] Accordingly, stereotypes are both a cause and a manifestation of structural disadvantage and are considered one of the primary impediments to gender equality.[158] This connection is increasingly recognised in international human rights law. As noted by the former UN Special Rapporteur on Violence against Women, Yakin Ertürk, law has begun to move 'past symptoms of gender inequality that become manifest as distinct forms of violence to look at structural and ideological causes that underlie the problem beyond the injury caused'.[159] This also indicates a shift from viewing stereotypes as an issue of individual mentality to a source of structural discrimination. As a consequence, international human rights law establishes obligations not only to prevent, investigate and punish specific incidents of violence against women, but also to identify the underlying structures and ideologies that manifest themselves as harmful acts. Accordingly, ICTs may be harmful not only through interpersonal violations but also through the online representation of a particular group. While certain material, such as harmful pornography and sexist hate speech, may affirm and exacerbate gender stereotypes *per se* even if no individual woman is harmed, accumulated

[155] *M. C. v Bulgaria* (ECtHR).

[156] *Volodina v Russia* (ECtHR), para. 132.

[157] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006), para. 73.

[158] Art. 5 of CEDAW. *See* also Cook and Cusack (2010), p. 104; Holtmaat (2012), p. 105; Ertürk (2004), p. 7.

[159] UN Special Rapporteur on violence against women, its causes and consequences, Yakin Ertürk, '15 Years of the United Nations Special Rapporteur on Violence against Women (1994–2009) – A Critical Review', UN Doc. A/HRC/11/6/Add.5 (27 May 2009), para. 87.

individual instances of online gender-based violence also contribute to its social anchoring. State obligations to eradicate gender stereotypes are thus relevant in both regards.

In analysing the role of ICTs in upholding and cultivating gender roles, several provisions in international human rights law are relevant. Most central is Article 5 (a) of the CEDAW, which obliges states to 'take all appropriate measures... to modify social and cultural patterns of men and women with a view to achieving the elimination of prejudices, customs and practices based on the inferiority or superiority of either sex or on stereotyped roles for men and women'. This obligation is an expression of equality as transformation,[160] that is, a means of effecting structural change by demanding the dismantling of gender stereotypes and is distinctive in that it does not solely focus on legislative obligations. Similar obligations are included in regional women's rights conventions.[161] Gender stereotyping has also been recognised as an aspect of discrimination in case law of regional human rights law courts.[162] Differences in treatment on the basis of gender stereotypes, or effects that encourage stereotyping, may thus constitute discrimination. Gender stereotypes have also been identified within the scope of other rights, such as the right to privacy.[163]

Although obligations to combat gender stereotyping are becoming increasingly concrete, the understanding of what constitutes "gender stereotyping" from the perspective of these conventions is, nevertheless, unclear.[164] No definition of gender stereotyping is included in human rights law treaties, although several bodies of the UN have approached stereotyping as a process of 'ascribing to an individual general attributes, characteristics, or roles by reason only of his or her apparent membership in a particular group'.[165] The CoE has in relation to *gender* stereotyping defined it as '...preconceived ideas whereby males and females are arbitrarily assigned

[160] Holtmaat (2004), p. xii.

[161] Art. 12 (1) of the Istanbul Convention; Art. 8 (b) of the Belém do Pará Convention; Art. 2 (2) of the Maputo Protocol.

[162] *Konstantin Markin v Russia* (2013) 56 EHRR 8; *Case of González* et al. *(Cotton Field) v Mexico* (IACtHR), para. 401.

[163] *See,* for example, *Y v Slovenia* (2016) 62 EHRR 3; *J. L. v Italy* App no 5671/16 (ECtHR, 27 May 2021) (the manner in which domestic court proceedings were conducted in a case of sexual violence).

[164] In contrast, *see* Cook and Cusack (2010) which delineates "sex", "sexual" and "sex role" stereotyping. Sex stereotyping involves the promotion of perceived physical and biological differences between men and women, such as strength, cognitive or emotional differences, beyond gender role implications. Meanwhile, sexual stereotypes centre on the sexual interaction between men and women. It pertains to issues such as sexual initiation, sexual assault and objectification, mainly concerning dominant male sexuality and passive, modest female sexuality. Sex role stereotypes describe the roles and behaviour ascribed to men and women as a result of physical, social and cultural differences.

[165] Cook and Cusack (2010), p. 1. This definition has also been adopted by UN bodies. *See*, for example, UN OHCHR, 'Gender Stereotyping as a Human Rights Violation' (October 2013), p. 8.

characteristics and roles determined and limited by their sex'.[166] Furthermore, the IACtHR has held that it 'refers to a preconception of personal attributes, characteristics or roles that correspond or should correspond to either men or women'.[167]

Meanwhile, the CEDAW Committee has not defined "stereotypes" or the "inferiority or superiority of either sex" in relation to Article 5 but requires the eradication of "harmful" gender stereotypes and "wrongful" gender stereotyping. What this involves is not clear. "Harmful" does not necessarily entail that the stereotype is negative. It can also appear to be benign, such as women being characterised as nurturing, or involving positive comments on their physical appearance. Nevertheless, "harmful" stereotyping must, at a minimum, pertain to acts or speech with such consequences as the denial of an individual's dignity or a benefit; the imposition of a burden; the degradation of women or the unequal distribution of public goods.[168] From this viewpoint, a stereotype is harmful if it nullifies the enjoyment of human rights and freedoms. At the same time, it appears from concluding observations and general recommendations that the CEDAW Committee mainly associates harmful stereotyping with depictions of women as sex objects or homemakers, and women as submissive and associated with the private sphere, limiting women's participation in the public sphere.[169] It is thus in practice a more narrow approach.

Furthermore, in CEDAW General Recommendations No. 19 and No. 35, a direct link between negative stereotyping and violence against women is recognised. The CEDAW Committee considers that gender-based violence stems from such factors as '. . .the ideology of men's entitlement and privilege over women, social norms regarding masculinity, the need to assert male control or power, enforce gender roles, or prevent, discourage or punish what is considered to be unacceptable female behaviour'.[170] For example, as will be discussed further in Sect. 4.5, the CEDAW

[166] CoE, 'Gender Equality Strategy 2014–2017' (February 2014), p. 9. The CoE has also developed standards in relation to sexism, which is understood as 'the supposition, belief or assertion that one sex is superior to the other'. *See* Council of Europe, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 3.

[167] *González* et al. *(Cotton Field) v Mexico* (IACtHR), para. 401.

[168] Cook and Cusack (2010), p. 59.

[169] CEDAW, 'General Recommendation No. 21: Equality in Marriage and Family Relations' (12 April 1994) UN Doc. A/49/38. Work in the home is generally seen as inferior and thus less valued. In relation to Indonesia, it considered that '. . .the existence of cultural attitudes that confine women to the role of mothers and housewives presents a great obstacle to the advancement of women'. *See* CEDAW, 'Concluding Observations on Indonesia (14 May 1998) UN Doc. A/53/38, CEDAW/C/SR. 377, para. 282. *See* also CEDAW, 'Concluding Observations on Armenia' (12 August 1997) UN Doc. A/52/38/Rev.1 Part II, 78, para. 65. In a report on Germany, it expressed concern '. . .that women are sometimes depicted by the media and in advertising as sex objects and in traditional roles'. *See* CEDAW, 'Concluding Observations on Germany' (18 March 2004) UN Doc. A/59/38, CEDAW/C/SR.640 and 641, para. 384.

[170] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 19. General Recommendation No. 19 holds that '[t]raditional attitudes by which women are regarded as subordinate to men or as having stereotyped roles perpetuate widespread practices involving violence or coercion, such as family violence and abuse. . . [s]uch prejudices and practices

Committee and the Beijing Declaration and Platform for Action have linked the prevalence of pornography and sexualised images of women and girls to increased levels of gender-based violence.[171] Additionally, by not responding effectively to violence against women, states enforce the stereotype that women are inferior, through failing to consider this type of violence a serious crime. This has also been affirmed in the case law of regional human rights law courts.[172]

In terms of Article 5 of the CEDAW, state obligations encompass both acts and omissions by states and non-state actors. The Article thus has a twofold implication: states must adopt policies that aim to eliminate stereotyped images of men and women in addition to considering the gendered construct and indirectly discriminatory effects of laws or policies.[173] The measures required by the CEDAW Committee appear in practice to mainly involve state-run information campaigns, education on gender roles and taking steps to influence the media and advertising.[174] This includes the promotion of media literacy skills in school curricula, including analysis of the construction of gender roles and the nature of consent and coercion. It also involves general awareness-raising programmes and sensitisation of judges, lawyers and law enforcement personnel, in order to dismantle the commonly held victim-blaming beliefs that make women responsible for their own safety and for the violence they suffer, an approach also taken by certain regional institutions.[175]

Of relevance to ICT-related harm, as mentioned above, the need for women's participation in the development of new technology has been affirmed as a means of counteracting harmful online content. State parties to the CEDAW are also obliged to take measures to influence third parties, including the media and ICTs, in

---

may justify gender-based violence as a form of protection or control of women'. *See* CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 11.

[171] The Beijing Declaration and Platform for Action (Platform for Action), para. 118, holds that images of violence against women in the media, such as rape, sexual slavery, or as sex objects—as in pornography—contribute to the prevalence of such violence, '. . .adversely influencing the community at large'; CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 12.

[172] *González* et al. *(Cotton Field) v Mexico* (IACtHR), para. 401; *Veliz Franco* et al. *v Guatemala* (preliminary objections, merits, reparations and costs), IACtHR Series C No 277 (19 May 2014), para. 213; *Volodina v Russia* (ECtHR), para. 132; *Opuz v Turkey* (ECtHR), para. 198.

[173] Holtmaat (2004), p. xii. The more concrete content of Art. 5 has been developed in general recommendations on specific topics, such as violence against women (*see* Gen. Rec. No. 12, 19 and 35).

[174] Holtmaat (2004), p. 71.

[175] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 34 (2 and b). *See* also CEDAW, 'General Recommendation No. 36 on the Right of Girls and Women to Education' (16 November 2017) UN Doc. CEDAW/C/GC/36, para. 72, CEDAW, 'Concluding Observations on Lithuania' (2000), UN Doc. A/55/38, CEDAW/C/SR. 472, 473 and 480, para. 139; *González* et al. *(Cotton Field) v Mexico* (IACtHR), para. 531. Art. 17(2) of the Lanzarote Convention also provides that '[p]arties shall develop and promote, in co-operation with private sector actors, skills among children, parents and educators on how to deal with the information and communications environment that provides access to degrading content of a sexual or violent nature which might be harmful'.

preventing gender stereotypes. Nevertheless, this obligation is mainly construed as involving the creation or strengthening of self-regulatory mechanisms to address harmful content through their services and platforms.[176] This includes representations of male and female roles in pornography and negative images that inspire violence against women.[177] As such, obligations are not discussed in terms of adopting liability regimes but rather affirm the *status quo* of self-regulation. At the same time, in a concluding observation on Sweden, the Committee noted the prevalence of stereotyped and sexualised images of women in the media, arguing that self-regulation was not sufficient to address the issue.[178] It should also be noted that the EU Directive on Audiovisual Media Services obliges states to restrict certain media outlets containing speech categorised as sexist hate speech, and the proposed Digital Services Act (DSA) places obligations on Very Large Online Platforms (VLOPs) to address harmful material.[179] Nevertheless, as will be discussed in Sect. 3.4, state obligations to establish secondary liability regimes for Internet intermediaries and media publishers are limited in relation to gender-based harm, in both international human rights law and EU law. This is even less so in relation to content categorised as harmful but not illegal, the former understood as encompassing gender stereotypes. For example, regulating such liability has been affirmed as a state obligation by the European Convention on the Protection of Human Rights and Fundamental Freedoms (ECHR) solely in relation to hate speech, as delineated by the Court, which does not encompass sexist speech.[180] As the Internet in practice is mainly governed by corporations, state obligations to regulate the content in this sphere is thus rather limited.

The CEDAW Committee is also increasingly more specific in linking Article 5 to legislation, by combining Article 2(f) and Article 5(a).[181] For example, in relation to

[176] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 37 (a). In relation to such regulatory mechanisms, states must also establish and/or strengthen the capacity of national human rights institutions to monitor or consider complaints regarding gender-discriminatory images in the media (ibid para. 37 (c)). *See* also a similar approach by UN Women in cooperation with ESCAP, UNDP, UNFPA, UNICEF and WHO, 'Report of the Expert Meeting on Prevention of Violence against Women and Girls' (2012) EGM/PP/2012/Report, paras. 30, 134.

[177] Holtmaat (2004), p. xii: CEDAW, 'Concluding Observations on the Combined Eighth and Ninth Periodic Reports of Sweden' (10 March 2016) UN Doc. CEDAW/C/SWE/CO/8–9, para. 24 (c).

[178] CEDAW, 'Concluding Observations on the Combined Eighth and Ninth Periodic Reports of Sweden' (10 March 2016) UN Doc. CEDAW/C/SWE/CO/8–9, para. 24 (b).

[179] Audiovisual Media Services Directive (2018), Art. 6 (1) (a) and Art. 9 (c) (ii); Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020), Art. 26 and Art. 27. The CoE has also adopted principles for media organisations in the elimination of sexism. *See* CoE, 'Recommendation CM/Rec(2013)1 of the Committee of Ministers to Member States on Gender Equality and Media (Adopted by the Committee of Ministers on 10 July 2013 at the 1176th meeting of the Ministers' Deputies), para. 4.

[180] *Delfi v Estonia* (ECtHR).

[181] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 26. For an overview, *see* Holtmaat (2004), p. 72.

gender-based violence, states must examine gender-neutral laws and policies and ensure that they 'do not create or perpetuate existing inequalities and repeal or modify them if they do so'.[182] As noted by the UN Secretary-General, states have '. . .shaped cultural and social norms through laws and policies that incorporated existing gender relations of power or modified them to respond to State-centred goals. . .'.[183] Seemingly gender-neutral legislation may thus stem from and perpetuate gender stereotypes. Regional human rights law treaties on violence against women and women's rights similarly call for the modification of cultural patterns affirming gender stereotypes, including tolerance of violence against women, affecting legislative obligations.[184] Obligations *vis-à-vis* laws upholding stereotypes have also been formulated in case law by regional human rights law courts in relation to the right to privacy and the prohibition on inhuman or degrading treatment.[185]

Whether obligations arise for states to restrict gender stereotypes expressed by individuals on the Internet, either through civil or criminal law, depends on the content and context of the stereotypes expressed, as these may be related to a variety of offences. It may include domestic laws on anti-discrimination, hate speech, harassment or violence against women, that is, involving both individual victims and women as a group.[186] More specific obligations have thus been developed by regional human rights law courts in relation to select offences and materials, although not generally involving or highlighting the aspect of gender stereotyping. Positive obligations associated with a particular violation thus arise, be it the prohibition on sexual violence or hate speech. The ECtHR has, for instance, accepted the criminal prosecution of individuals associated with websites publishing obscene pornography on the Internet.[187]

2.2.2.4.4 Online Gender-Based Violations

2.2.2.4.4.1 International Human Rights Law

A range of gender-based harm adversely affecting women is widespread on the Internet. These can be divided into two broad categories, although overlapping in certain instances: harm generated through *speech* and information-sharing on the one hand, and *physical* acts committed through a digital forum on the other. Meanwhile,

[182] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 32.

[183] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006), para. 101.

[184] Art. 2 (2) of the Maputo Protocol; Art. 12 (1) and Art. (2) of the Belém do Pará Convention.

[185] *T. M. and C. M. v Moldova* App no 26608/11 (ECtHR, 28 January 2014); *M.C.v Bulgaria* (ECtHR).

[186] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 3.

[187] *Perrin v the United Kingdom* App no 5446/03 (ECtHR, 18 October 2005).

a definition of the concept of online gender-based violations is lacking in international treaties and case law. Various terminology is used by international organisations and scholars, including "online or cyber gender-based violence",[188] "online or cyber violence against women",[189] "technology-facilitated gender-based violence"[190] and "the digital dimension of violence against women".[191] Concepts have also been developed concerning particular online offences, such as "image-based sexual abuse" or "technology-facilitated sexual violence".[192] There are thus variations in terms of employing a gender neutral approach and/or a focus on women, and either exclusively addressing harm on the Internet or various forms of technology. Whereas online violence concerns activities and materials on the Internet, technology-facilitated violence encompasses activities performed through the use of technology and communication equipment, including smartphones, cameras and GPS.[193] As noted by international and regional organisations, the lack of a comprehensive typology stems from the diversity of aims and perspectives of different stakeholders.[194] This fragmentation impedes the development of effective prevention strategies, including precise and harmonised definitions in domestic law.

Beyond the variations of these concepts, they reflect a similar view on technology as an enabling tool, rather than the main impetus for violence. Although offences

[188] European Parliament resolution of 14 December 2021 with recommendations to the Commission on combating gender-based violence: cyberviolence (2020/2035(INL)); European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[189] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 35; UN HRC, 'Resolution adopted by the Human Rights Council on 5 July 2018: Accelerating Efforts to Eliminate Violence against Women and Girls: Preventing and Responding to Violence against Women and Girls in Digital Contexts' (17 July 2018) UN Doc. A/HRC/RES/38/5; European Institute for Gender Equality (EIGE), 'Cyber Violence against Women and Girls' (2017), p. 2; UNHRC, '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).

[190] UNFPA, 'Technology-Facilitated Gender-Based Violence: Making All Spaces Safe' (2021); Dunn for Centre for International Governance Innovation, 'Technology-Facilitated Gender-Based Violence: An Overview' (2020) <https://www.jstor.org/stable/pdf/resrep27513.1.pdf> Accessed 29 March 2022.

[191] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women'.

[192] Henry et al. (2020) and McGlynn and Rackley (2017).

[193] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 23.

[194] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 63; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 28; OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', Prepared by the General Secretariat of the OAS (2021) OEA/Ser.D/XXV.25, p. 27.

perpetrated through new technologies, particularly the Internet, in certain instances are considered a distinct category—"cybercrimes"—ICT-related violations are thus in the main viewed as new forms of pre-existing oppressions.[195] According to the UN Special Rapporteur on the Freedom of Expression, in the same manner as a telephone can be used to conspire to commit as well as report crimes, so does the Internet provide a platform for a variety of activities.[196] Accordingly, the CEDAW Committee and the UN Special Rapporteur on Violence against Women have emphasised that online gender-based violence stems from the same root causes as gendered offences in general and must be addressed in the broader context of eliminating discrimination against women.[197] A similar stance has been taken by the UN General Assembly,[198] the CoE,[199] the EU[200] and the OAS.[201] This entails that the general framework on gender-based violence, including the CEDAW and the regional treaties on gender-based violence, applies. However, even if the causes and consequences of such violence are similar regardless of context, its expression varies.[202] Certain forms of violence may be exacerbated and new forms arise in tandem with technological changes, which must inform the interpretation of the scope of rights and state obligations. As such, '[v]iolence against women is both universal and particular'.[203] Thus, although the well-established concept of

[195] Cybercrime: Council of Europe, 'Group of Specialists on the Impact of the use of New Information Technologies on Trafficking in Human Beings for the Purpose of Sexual Exploitation: Final Report' (16 September 2003) EG-S-NT (2002) 9 Rev., p. 7. New form of pre-existing violence: UNHRC, '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), para. 20.

[196] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 2.

[197] UNHRC, '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), para. 21; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 20.

[198] UNGA, 'Promotion of the Declaration on the Right and Responsibility of Individuals, Groups and Organs of Society to Promote and Protect Universally Recognized Human Rights and Fundamental Freedoms: Protecting Women Human rights Defenders' (30 January 2014) UN Doc. A/RES/68/181.

[199] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 10.

[200] European Parliament resolution of 14 December 2021 with recommendations to the Commission on combating gender-based violence: cyberviolence (2020/2035(INL)); European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', Art. 37 (5).

[201] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 5.

[202] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006), para. 71.

[203] ibid., para. 69.

gender-based violence applies online, which forms of content and conduct are included must, to an extent, be addressed in a contextual manner.

The general approach at the international level is to adhere to the formula of General Recommendation No. 19, applied to the online context. That is, given that the form of offences is affected by social, technological and political developments, an exhaustive list of online gender-based violations is avoided for the benefit of focusing on the nature, purpose and effect of harm. The UN Special Rapporteur on Violence against Women has accordingly considered that gender-based violence against women through new technologies is violence which is '. . .committed, assisted or aggravated in part or fully by the use of ICT, such as mobile phones and smartphones, the Internet, social media platforms or email, against a woman because she is a woman, or affects women disproportionately'.[204] A similar delineation has been made by other UN bodies and the CoE.[205] For example, offences *committed* through the use of ICTs include image-based sexual abuse, whereas ICTs may *facilitate* stalking and *aggravate* defamation. This is further addressed in Sect. 2.3.

International organisations have also, in a general manner, categorised certain Internet offences as gender-based and as particularly affecting women and girls. Frequently, this involves cyber stalking, image-based sexual abuse, harassment, unsolicited pornography, sextortion, rape and death threats, doxing, and electronically enabled trafficking.[206] Online abuse, bullying and pornographic and sexist websites have also been addressed as impediments to gender equality, since women are both disproportionately targeted and suffer distinct forms of harm.[207] Similarly,

[204] UNHRC, '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), para. 23.

[205] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 35; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 29.

[206] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 6; UNGA, 'Intensification of Efforts to Prevent and Eliminate all Forms of Violence against Women and Girls: Sexual Harassment' (14 November 2018) UN Doc. A/C.3/73/L.21/Rev.1 (stalking, death threats and threats of sexual and gender-based violence, trolling, cyberbullying and other forms of cyberharassment, including unwanted verbal or non-verbal conduct of a sexual nature); UNGA, 'Promotion of the Declaration on the Right and Responsibility of Individuals, Groups and Organs of Society to Promote and Protect Universally Recognized Human Rights and Fundamental Freedoms: Protecting Women Human rights Defenders' (30 January 2014) UN Doc. A/RES/68/181 (online harassment, cyberstalking, violation of privacy, censorship and hacking of email accounts, mobile phones and other electronic devices); CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 33 (non-consensual image or video sharing, coercion and threats, including rape threats, sexualised bullying and other forms of intimidation, online sexual harassment, impersonation, online stalking or stalking via the Internet of Things as well as psychological abuse and economic harm perpetrated via digital means against women).

[207] UNHRC, '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'

the Council of the European Union considers 'emerging forms of violations' such as online harassment, sexual abuse instigated or facilitated through the use of ICTs, stalking, and bullying as gender based-violence against women.[208] In *Buturuga v Romania*, the accessing of the private social media accounts of the victim without consent in the context of domestic violence was recognised by the ECtHR as a form of online violence against women.[209] While these forms of violence in most instances occur between private individuals, it also encompasses institutional violence, where a state actor commits such acts, for example, to further a particular ideological agenda.[210]

Furthermore, certain gender-based offences transpire both online and offline, such as sexual violence and sexual harassment. Such violations have been categorised by various international courts and committees as having gendered causes and consequences *per se*.[211] Other forms of online harm are novel as a result of technological developments, such as the non-consensual distribution of intimate images online. Whether such technology-facilitated harm is encompassed within existing definitions of, for example, sexual violence, is less clear. Meanwhile, certain offences, such as defamation, are not considered inherently gendered but are in certain instances perpetrated in a gender-based manner and have gendered effects on the Internet. Additionally, whether the regulation of particular forms of speech, including pornography and sexist hate speech, is considered part of international human rights law is contested *per se*, but must be considered anew in the online context. These latter types of speech are mainly addressed in relation to gender stereotyping.

2.2.2.4.4.2 Empirical Studies and Theories

Beyond this legal framework, the gender-based implications of offences on the Internet are made apparent through empirical studies indicating the *prevalence* and *nature* of online harm. Due to the relative novelty of the issue, international surveys mapping the occurrence of online offences are still few in number. At a general level, the UN Working Group on Broadband and Gender noted in 2015 that 73% of

(18 June 2018), paras. 20, 25, 28. In its 'General Recommendation No. 36 on the Right of Girls and Women to Education' (16 November 2017), para. 70, the CEDAW Committee also acknowledges how girls are affected by cyberbullying.

[208] Council of the European Union, 'Council Conclusions: Preventing and Combating all forms of Violence against Women and Girls, Including Female Genital Mutilation' (2014), para. 14.

[209] *Buturuga v Romania* (ECtHR), para. 74.

[210] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 102.

[211] Art. 2 of the Belém do Pará Convention; CEDAW, 'General Recommendation No. 19 on Violence Against Women', paras. 6 and 17; Art. 3 (a), 36 and 40 of the Istanbul Convention; ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', para. 57.

women had been exposed to or had experienced some form of online violence.[212] More specifically, a 2014 EU survey demonstrated that 11% of women had been subjected to online harassment,[213] and 25% in a CoE study indicated that they had been exposed to online sexual harassment.[214] In a survey commissioned by Amnesty International in 2017 on online abuse against women, 46% of women who had experienced online abuse or harassment indicated that it was misogynistic or sexist in nature.[215] Between one-fifth and one-quarter of women who had been subjected to abuse or harassment held that it had included threats of physical or sexual assault.[216] In 26% of these instances, personal or identifying details of them had been shared online, also known as "doxing".[217] Female names of Internet users attract a substantially higher degree of malicious messages than male names.[218] As a consequence, women to a greater extent than men become non-contributing observers, or retreat to women-only lists.[219] It is also common for women to adopt gender-disguising names or to downplay stereotypical female attributes on the Internet,[220] thus in effect limiting the visible female presence on the Internet.

The surveys indicate that online gender-based offences primarily affect girls and young women, as they are more likely to use social networking sites and instant messaging.[221] The risk of victimisation for women in the 18–29 age group is twice as high as for women between 40 and 49, and three times as likely as for women

[212] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call' (23 October 2015), p. 2. *See* also an IPSOS Mori poll in 2017, commissioned by Amnesty International, which looked at the experiences of women between the ages of 18 and 55 in eight countries, according to which 23% of women held that they had experienced online abuse or harassment at least once, ranging from 16% in Italy to 33% in the US, <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022.

[213] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 95. The study is based on interviews with 42,000 women across 28 member states.

[214] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 7. In a survey by Powell and Henry, 10.1% of women in the UK and 11.7% of women in Australia responded that they had an unwanted sexual experience with someone they met online, with 21.1% of women in the 20–24 age group. *See* Powell and Henry (2017), p. 85.

[215] IPSOS Mori poll: <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022.

[216] Also, in the survey by Powell and Henry, approximately 9% indicated that they had been exposed to threats of rape online, varying from credible threats against individual women to those directed at women as a group. *See* Powell and Henry (2017), p. 93.

[217] IPSOS Mori poll: <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022.

[218] Citron (2014), p. 14.

[219] Herring et al. (1995), p. 69.

[220] Citron (2010), p. 40.

[221] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), pp. 87, 93, 95. The risk of young women experiencing such violations was twice as high as for women between 40–49; UN Women in cooperation with

between 50 and 59.[222] Furthermore, cyber mobs more frequently target lesbian and/or non-white women,[223] and especially young women who are visible in the media, feminist or women's rights defenders.[224] Women voicing their opinions in public debate are disproportionately harassed, in part due to the perceived transgression of traditional gender roles.[225] Cyber harassment is also more commonly experienced by women with a university degree and in the highest occupational groups.[226] For example, in a CoE study on harassment against female members of parliament in 45 European countries, 58% replied that they had been the target of online sexist attacks on social networks, including images or comments.[227] 47% of the respondents also reported receiving death threats, threats of rape and beatings, involving them, their children or their families.[228]

The UN Human Rights Council has also noted the particular vulnerability of women journalists, media workers, public officials or others engaging in public debate.[229] For instance, surveys indicate that between 44% and 73% of female journalists have been harassed online.[230] In a report by the British newspaper *The Guardian*, evaluating its online articles and reader comments, hateful remarks were more commonly found in response to articles written by female journalists and/or articles concerning feminism or sexual violence.[231] Additionally, online harassment is more common in countries with high rates of Internet access, such as Denmark and

---

ESCAP, UNDP, UNFPA, UNICEF and WHO, 'Report of the Expert Meeting on Prevention of Violence against Women and Girls' (2012), para. 30.

[222] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 105.

[223] Citron (2010), p. 32.

[224] CoE, 'Seminar Combating Sexist Hate Speech: Report', 10–12 February, EYC, Strasbourg, p. 19; UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 17; European Parliament resolution of 11 September 2018 on measures to prevent and combat mobbing and sexual harassment at workplace, in public spaces, and political life in the EU (2018/2055(INI)), para. AE.

[225] Megarry (2014), p. 48.

[226] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 96.

[227] CoE Parliamentary Assembly, 'Sexism, Harassment and Violence against Women in Parliaments in Europe', Issues Brief (October 2018), p. 6.

[228] ibid., p. 7.

[229] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017) UN Doc. A/HRC/35/9, para. 36. *See* also UNHRC, '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), para. 28.

[230] UNESCO, 'Online Violence against Women Journalists: A Global Snapshot of Incidence and Impacts' (2020) <https://unesdoc.unesco.org/ark:/48223/pf0000375136> Accessed 10 March 2022. *See* also an International Federation of Journalists survey, press release on 24 November 2017 <https://www.ifj.org/media-centre/news/detail/category/press-releases/article/ifj-survey-one-in-two-women-journalists-suffer-gender-based-violence-at-work.html> Accessed 10 March 2022.

[231] Gardiner (2018).

Sweden.[232] These surveys thus indicate that an intersectional approach is important, bearing in mind such identity characteristics as sex, ethnicity, sexual orientation, career, and age.[233] Although perpetrators often are anonymous,[234] certain studies indicate that these are primarily men, whereas others demonstrate a varied demographic.[235]

In terms of the nature of the violations, much defamation, harassment and other harmful material on the Internet involve the objectification of women, that is, treating women as objects for use and abuse by men.[236] Female objectification mainly entails that femininity is linked to women's visual appearance, with women assessed and treated as bodies or body parts.[237] It includes comments on physical attributes, as either objectionable or sexualised, where unattractiveness is a central insult.[238] It thus involves both hostile and benevolent sexist gender biases.[239] For example, a study of the comment section of several large social media platforms, including YouTube, indicated that female content creators received more negative comments, such as sexually aggressive and racist speech, but also a higher degree of seemingly benign comments on their physical appearance.[240] Insulting speech against women online often involves also other types of gender stereotypes, such as a lack of intellectual capacity. In contrast, men are frequently insulted in relation to their professional capacity.[241] Women are thus in this regard construed as objects, whereas men are treated as subjects.[242]

Sexual harassment is the most common way of harming women online, including the receipt of unsolicited nude images and the non-consensual publication of intimate photographs of the victim.[243] The UN Special Rapporteur on Violence against Women notes that online threats against women are often misogynistic and

[232] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 104.

[233] UNHRC, '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), para. 28.

[234] *See* CoE Parliamentary Assembly, 'Sexism, Harassment and Violence against Women in Parliaments in Europe', p. 6 (67%); IPSOS Mori poll: <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022 (59%).

[235] Powell and Henry (2017), p. 249. *See* also an overview in NIKK, 'Hat och Hot på Nätet: En Kartläggning av. den Rättsliga Regleringen i Norden Utifrån ett Jämställdhetsperspektiv' (2017), <https://nikk.no/wp-content/uploads/2019/10/2017-Hat-och-hot-pa-natet.pdf> Accessed 15 February 2022, p. 26.

[236] Nussbaum (2010), p. 68.

[237] Marganski (2018), p. 14.

[238] Megarry (2014), p. 49.

[239] Döring and Mohseni (2020), p. 66.

[240] ibid., p. 64.

[241] Bladini (2020), p. 23.

[242] ibid., p. 4.

[243] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 20.

sexualised.[244] For example, although both female and male journalists on many occasions are subjected to online threats, the nature of the threats is often different, as they frequently involve sexualised threats when directed against women, in contrast to threats of assault against men.[245] Similarly, defamation, which is generally perceived as a gender-neutral offence, has gendered causes and effects *vis-à-vis* women. Defamatory statements on public fora involving women often centre on sexuality, for example, with claims of promiscuity or STDs.[246] When women's behaviour is considered to deviate from gender norms on submissiveness, harassment also tends to centre on positioning them as sex objects.[247]

Gender stereotypes play a role in the choice of speech. The harasser is aware of the gendered effects of speech in that women, as opposed to men, generally will be perceived as promiscuous. This is, for example, apparent in relation to image-based sexual abuse. Female sexuality has traditionally been associated with passivity, to be exercised within marriage as a service to a husband and for the purpose of procreation. Arguably, female sexuality is still dependent on male initiation.[248] As a result, women who take such photographs or risks are deemed responsible for their own harm. The shame and reputational harm thus to a degree stem from the stigma attached to female bodies and sexuality.[249] Accordingly, '[i]t is the societal reactions that place a moral judgment on women, their sexuality, and in particular their sexual agency that make image-based sexual abuse such an effective method of inflicting harm'.[250] This may also pertain to the exposure of male sexuality. However, in such instances, the humiliating aspect is to a higher degree relative to the particular content of a photograph (e.g., gay sex) or the social standing of the victim, whereas for women nude images are generally deemed reprehensible *per se*, as they transgress norms on female sexuality.[251]

Even the distribution of sexual assault images online in certain instances generates hostile comments against female victims.[252] In several cases at the domestic

[244] UNHRC, '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), para. 29.

[245] NIKK, 'Hat och Hot på Nätet: En Kartläggning av. den Rättsliga Regleringen i Norden Utifrån ett Jämställdhetsperspektiv', p. 24; Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, 'Women Journalists and Freedom of Expression: Discrimination and Gender-Based Violence Faced by Women Journalists in the Exercise of their Profession' (31 October 2018) OEA/SER.L/V/II, CIDH/RELE/INF.20/18, para. 45; UNESCO, 'World Trends in Freedom of Expression and Media Development: 2017/2018 Global Report', p. 157.

[246] Franks (2012), p. 683; Citron (2009), p. 389.

[247] Megarry (2014), p. 50.

[248] Patton (2015), pp. 414, 419.

[249] Suzor et al. (2017), p. 1061.

[250] Powell and Henry (2017), p. 242.

[251] ibid., p. 243.

[252] *See* discussion in Citron (2014), p. 114.

level where such videos have been disseminated, subsequent online cyberbullying has centred on the degradation of the female victim by, for example, chastising the victim's appearance and behaviour and, in certain instances, included encouragement of the rapists.[253] Women and girls who transgress norms on aggressive male sexuality and female modesty and passivity—regardless of the non-consensual nature of the act—are punished socially, with the female victim held responsible for failing to avoid sexual violence.[254]

Through pornography women are also commonly portrayed as sexual commodities for the consumption of the viewer.[255] The distribution and consumption of pornography has escalated immensely, with certain studies indicating that the Internet not only has facilitated its accessibility but also affected its content, which is increasingly hardcore and violent towards women. [256] The UN Working Group on Broadband and Gender notes that 30% of all Internet traffic constitutes pornography.[257] 88% of top rated pornographic scenes on the Internet contain aggressive acts and 94% of such acts are directed towards a woman.[258] As noted, a correlation between the stereotyped images of women and gender-based harm in cyberspace can be made. However, the link between pornography and gender-based violence is a contentious area and will be further discussed in Sect. 4.5.

The existing framework in international human rights law—categorising certain acts and content as gender-based and as affirming gender stereotypes—as well as empirical studies, thus demonstrate that the forms of online conduct and content analysed in this book are encompassed, requiring positive measures by states to protect individuals and to regulate harmful material.

2.3 Gendering Features of the Internet

2.3.1 Introduction

While the previous sections affirmed that international human rights law provisions apply equally online, including the obligation to protect individuals against

[253] *See* news article on cases in the US (Audrie Pott) and Canada (Rehtaeh Parsons): Washington Post, 'Audrie Pott: Sexual Assault, Cyberbullying and Suicide' (16 April 2013) <https://www.washingtonpost.com/blogs/therootdc/post/audrie-pott-rape-case-sexual-assault-cyberbullying-and-suicide/2013/04/16/d00eb450-a6b9-11e2-b029-8fb7e977ef71_blog.html> Accessed 10 March 2022.

[254] Dodge (2016), p. 71.

[255] Marganski (2018), p. 14.

[256] Bridges et al. (2007) and Vera-Gray et al. (2021).

[257] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 7.

[258] Bridges et al. (2007). As will be discussed in the section on pornography (Sect. 4.5), such statistics are also disputed. *See*, for example, Shor and Seida (2019).

gender-based violence and gender stereotyping, the context of the Internet is in many ways distinct. Accordingly, context-neutrality may be inimical to ensuring the objectives of international human rights law and the effective implementation of rights. That is, a consideration of the specific features of the Internet must be borne in mind in the application of rights pertinent to gender-based offences. This in turn may affect the interpretation of the scope of rights and obligations, including the balancing in conflicts of interests on the Internet.

In the following part, certain features of the Internet will be highlighted in order to explain the manner in which certain gender-based offences are facilitated or exacerbated online, and challenges that arise in effectively regulating this sphere. The gendered process in developing the Internet, its design, social norms and gaps in domestic regulation are offered as factors affecting the occurrence of gender-based harm. This is addressed through general theories on the constraints on online behaviour and the framework of cyberfeminism.

In order to explore the broader relationship between law and the Internet, Lawrence Lessig in his seminal work of *Code and Other Laws of Cyberspace* identified four modalities in regulating conduct both In Real Life (IRL) and online.[259] The constraints were categorised as law, architecture, norms and the market, although they frequently overlap. While Lessig's study primarily considered the necessity of either adopting context-specific legislation or maintaining technology-neutral regulation, the explanatory model of constraints on online conduct can also be employed to illustrate differences in how external restraints affect behaviour in this sphere. As will be made clear, the force of these constraints shifts online. The Internet may undermine or incapacitate certain forms of regulation, which affects the prevalence of harmful or illegal content. Since the focus of this book lies on law as a constraint, it is thus necessary to consider the indirect effects of law on other modalities of regulation and, in turn, how features of the Internet may affect the content of international human rights law.

Meanwhile, as the Internet expanded in the 1990s, cyberfeminist theories developed to explain the exacerbation of harmful *gendered* conduct and content online. Cyberfeminism encompasses a range of critical theories on the relationship between gender and digital culture[260] and analyses how the Internet operates, the power discourse that has an impact on its architecture and possibilities of using it for feminist objectives. Cyberfeminism further theorises the desirability or disadvantages of legal regulation of content on the Internet.[261] However, similar to feminist legal theories, cyberfeminist theories are multiple from an epistemological standpoint and in their approach to the relationship between men and women, with a core essentialist/constructivist divide, affecting the approach to gender and cyberspace. Many of the differences are positioned in binaries: a utopian/dystopian view of the utility of the Internet—as a sphere of liberation or oppression—and a focus on the

[259] Lessig (2006).

[260] Plant (1996) and Haraway (1991).

[261] Bailey and Telford (2007), p. 246.

virtual/physical as separate spheres.[262] These theories have also fluctuated as Internet architecture and use has developed. Such theories are nevertheless useful in analysing different aspects of law, gender and the Internet.

2.3.2 Constraints of User Behaviour

2.3.2.1 Architecture

As noted above, the Internet is in international human rights law viewed as a platform through which certain gender-based offences are *committed, facilitated* or *aggravated*. This accordingly considers how technical features affect the nature and prevalence of gender-based harm. From a general standpoint, Internet architecture—involving hardware, software, protocols and the connection medium—constrains human behaviour by setting limits for possible conduct, what can be expressed and how it is regulated through law. For example, it regulates the means of protecting privacy and censoring speech. It also limits the identification of individuals and the content of data. Additionally, its design affects other modalities of constraint, including market control and social relationships and norms.[263]

Meanwhile, how Internet architecture affects *gender* relations has been addressed in cyberfeminist theories. The premise of such theories is that a distinction between science and ideology cannot be made, as scientific knowledge is socially produced.[264] Gender-based online offences thus do not simply occur as a result of 'rational technical imperatives', that is, inevitable design choices that are abused by individual perpetrators.[265] While gender-based violence on the Internet stems from the same root causes as offline violence, technological infrastructure also exists in that paradigm of gender inequality. Such hierarchies may be re-inscribed or exacerbated by Internet architecture, be it by design or effect.[266]

According to Judy Wajcman, if technology is approached as neutral, with risks of misuse, the policy decisions involved in its design and development will be overlooked.[267] Developers define the research agenda by setting priorities and preferences and in many instances select particular technologies over others.[268] The view of technology as '. . .an external, autonomous force exerting an influence on society, narrows the possibilities for democratic engagement with technology, by presenting a limited set of options: uncritical embracing of technological change,

[262] Schulte (2011), p. 734.
[263] Lessig (2006), p. 125.
[264] Wajcman (2004), p. 17.
[265] ibid., p. 34.
[266] Schulte (2011), p. 735; Bailey and Telford (2007), p. 246; Wajcman (2004), p. 19.
[267] Wajcman (2004), p. 23.
[268] Elkin-Koren and Salzberger (1999), p. 579.

defensive adaption to it, or simple rejection of it'.[269] On the contrary, '[h]ierarchies of sexual difference profoundly affect the design, development, diffusion and use of technologies'.[270] This does not mean that each design choice reflects an ideology, as certain external technological limitations may arise. Nevertheless, technical knowledge and workplaces are frequently male-dominated, entailing that the development of design, production and use of technologies, including the Internet, often has a male orientation.[271] For example, software selection tends to favour the interests of boys and men, with predominantly male game characters and features.[272] Furthermore, the libertarian ideology, likewise categorised as a masculine preference, influences technological design, for example, the end-to-end construction of the Internet and user anonymity.[273] ISPs in the main also espouse libertarian ideals, for example, by placing the responsibility for posted content on users.[274] The political—including gender relations—is thus coded into ICTs.

With the view that technology is a result of policy decisions and thus can be re-directed to ensure social justice goals, features of the Internet that contribute to gender-based offences must be identified. Currently, several aspects of Internet architecture create technical challenges in regulating this forum, in addition to affecting social norms, thereby producing a platform conducive to certain forms of gender-based harm. Accordingly, several international organisations and scholars emphasise the distinct characteristics of technology-facilitated offences. As noted above, although offences such as harassment, stalking, and threats online fall within the scope of the UN definition of violence against women, the term "cyber" is accentuated in order to highlight the ways the Internet 'exacerbates, magnifies or broadcasts the abuse'.[275] According to the CoE, new technologies are not harmful in themselves, but they provide efficient and often anonymous methods of harming

[269] Wajcman (2004), p. 33.

[270] ibid., p. vii.

[271] For example, early cyberfeminist Donna Haraway noted that the male dominance in coding, design and computer and Internet production has led to the masculinisation of computer networking. *See* Haraway (1991). *See* also UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 25.

[272] Meyer (1999), p. 313.

[273] Shepherd et al. (2015), p. 8.

[274] Vitis and Segrave (2017), p. 9.

[275] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber Violence Against Women and Girls: A World-Wide Wake-Up Call' (2015), p. 21. *See* also UNHRC, '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), para. 20. According to the UN Special Rapporteur, patriarchal patterns that result in gender-based violence are reproduced but also amplified and redefined through ICTs, in addition to new forms of violence emerging.

individuals.[276] For example, Donna Hughes, through her work on online human trafficking, notes that new technologies *facilitate* sexual exploitation in that they enable the anonymous buying, selling and exchanging of images through diversified means.[277] At the same time, much focus has recently been placed on the online environment as harmful, as opposed to other means and methods of communication, such as postal services or airlines, which may also enable such gender-based offences as harassment or human trafficking. This has raised the question of whether there is confusion between a technical tool and the culture that uses it to harm others.[278] Although this focus may be due to its global reach and the increasingly integral part of the Internet in the daily lives of individuals, certain technical features of the Internet, in contrast to other tools and media, substantially enhance the risk of gender-based offences.

In terms of the gendered design and effect of ICTs, certain electronic devices have harmful objectives, such as particular forms of spyware, marketed with the aim of spying on a partner, and voyeurism cameras, for non-consensually taking photos under another person's clothing, most frequently used on female victims.[279] Other ICTs may have neutral purposes though the design has gendered effects. For example, technology can be a means of perpetrating domestic violence, through controlling the whereabouts of a partner via geolocations on social media, withholding access to technology, or by harassing victims through publishing defamatory statements online.[280] Image-based sexual abuse has also emerged through the possibility of user uploads of photographs.

At a general level, the vast increase in user-generated information and social interactions online have profoundly transformed the type of offences that occur. Whereas the Web 1.0 (websites, emails, and dial-up Internet) occasioned a domestic and international focus on pre-existing crimes that were amplified via technology, such as cyber security offences, financial crime, terrorism, and child abuse images, the development of the Web 2.0 (Wi-Fi, broadband Internet, and social networking sites) has generated widespread technology-facilitated interpersonal violence.[281] A key factor in the facilitation of interpersonal harm is thus the architecture and business model of social media companies, maximizing content sharing and user interaction, discussed below in terms of market constraints.[282]

---

[276] Council of Europe, 'Group of Specialists on the Impact of the Use of New Information Technologies on Trafficking in Human Beings for the Purpose of Sexual Exploitation: Final Report' (2003), p. 107.

[277] Hughes (2002), p. 129. *See* also CEDAW, 'Concluding Observations on Denmark' (12 August 1997) UN Doc. A/52/38/Rev.1 Part I, 34 at para. 269.

[278] Kathleen Maltzahn for Association for Progressive Communications, 'Digital dangers: Information & Communications Technologies and Trafficking', <https://www.apc.org/sites/default/files/digital_dangers_EN_1_0.pdf> Accessed 10 March 2022, p. 2.

[279] Powell and Henry (2017), p. 40; Tene and Polonetsky (2014), p. 83.

[280] *Volodina v Russia* (ECtHR), para. 75; *Buturuga v Romania* (ECtHR), paras. 73–78.

[281] Vitis and Segrave (2017), p. 2.

[282] Suzor et al. (2019), p. 94.

The technical features of the Internet also affect the nature and degree of harm. This includes the omnipresence of the Internet, its universal accessibility and the amplifying effect it has on data published online, including the permanence and asynchronicity of information.[283] There is a variety of formats, through images, live webcameras, chats and videos. ICTs are also fast, easy, and low cost, reducing time and distance, which influences social relationships.[284] The automation of ICTs eases human work in relation to tasks, for example, copying and distributing material. Activities such as information cascades and Google bombs involve the widespread distribution of information. This may lead to degrading images, such as intimate photos, quickly reaching a wide audience. Furthermore, content is frequently algorithmically amplified. For example, activities such as "liking" posts on social media have the effect of further distributing and "pushing" content.[285] This, in turn, may enhance existing biases in AI. The harm may also be prolonged as images and videos of a sexual nature posted without the consent of the person in question are difficult to permanently erase in the digital sphere, potentially causing long-lasting re-victimisation.[286] This entails that the effects on, for instance, reputation—as an aspect of the right to privacy—is more damaging, which may influence assessments of harm.[287]

Additionally, architectural features, such as its end-to-end design and user anonymity, affect the regulation of liability, thus limiting the impact of domestic laws.[288] The primary feature identified as exacerbating online gender-based harm is in fact user anonymity, with networks either allowing or disallowing the identification of individuals, depending on the pocket of the Internet.[289] This not only impedes investigations in instances of interpersonal harm but also has an impact on the development of particular social norms. Additionally, this sphere is to a large extent governed by Internet intermediaries, website operators and online media publishers, which facilitate offences by transmitting data or providing the platforms for content. Meanwhile, the direct communication between users further decentralises the source of information and reduces the role of media publishers and intermediaries in distributing information, curtailing their ability to manage content.[290] This disperses state control of the Internet, all the while states remain the subjects of international human rights law, with positive obligations to prevent gender-based violence.

In terms of the means of controlling content, as will be explored in Sect. 3.4, the structure of the Internet allows for broader possibilities of *ex ante* regulation, which

[283] Franks (2012), p. 682.
[284] Fascendini and Fialová (2011), p. 13.
[285] Henry et al. (2020), p. 1833.
[286] Citron (2014), p. 4.
[287] Leiter (2010), p. 167.
[288] Benedek and Kettemann (2013), p. 45.
[289] Franks (2012), p. 693.
[290] Elkin-Koren and Salzberger (1999), p. 562.

may be positive from the viewpoint of preventing the publication of harmful speech. Nevertheless, given the implications for the freedom of expression, both EU law and international human rights law place constraints on the monitoring, censoring and blocking of materials, with obligations solely arising under limited circumstances.[291] Furthermore, the inability to effectively zone areas of the Internet and limit access to certain content makes harmful material, such as obscene pornography, more easily accessible, including to minors.[292] This has an impact on the scope of liability regimes, in terms of technical solutions in controlling content. However, although technical barriers, such as age verification tools, may be utilised, harmful content does not solely affect children. As will be discussed further in the book, although it is commonly noted in this regard that Internet users actively search for and access information, users may still be exposed to harmful material accidentally or through malware. It may also have a negative overall effect on gender equality. Given the profound effect of these design features on online content and conduct, the options may be for law to either adapt to technology, through the development of new laws or context-sensitive interpretations, or to regulate architecture.[293]

2.3.2.2 Online Social Norms and the Market

Not only law but also social norms constrain behaviour through the threat of *ex post* punishment. While law achieves this through penalties, social norms may cause adverse social consequences.[294] Norms are informal standards and constraints on human behaviour commonly produced through custom, adherence to organisational structures and ethics and do not necessarily correspond to legal regulation, although laws may reinforce desired normative behaviour.[295] Meanwhile, *cyber* norms are informal social standards of appropriate user behaviour in cyberspace.[296] For example, flaming and spamming transgress Net etiquette and users who engage in such behaviour may be rebuked.[297] There is a convergence effect between cyber norms and traditional social norms, with cyber norms either developing in coherence with or in contrast to non-digital norms.[298] There is also an increasing impact of cyber norms on regular social norms and it is anticipated that this will escalate as the

[291] Discussed in Sect. 3.4.3.3.

[292] Lessig (1999), p. 504.

[293] *See* further in Sect. 3.4.3.4.

[294] Lessig, 2006, p. 124.

[295] Major (2000), pp. 62, 64.

[296] ibid., p. 70. "Cyber norms" is also increasingly used as a term to describe voluntary codes of conduct of intermediaries and principles guiding state conduct. *See*, for example, UNGA, 'Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security: Note by Secretary-General' (22 July 2015) UN Doc. A/70/174.

[297] Spinello (2001), p. 139.

[298] Major (2000), p. 60. *See* also Schauer (1998), p. 563.

Internet becomes more integral to the lives of individuals and society at large. The upholding of a dichotomy of the virtual/physical may not only disregard the severity of the harm experienced but also the effects of online offences, which are viewed as insulated and not affecting social development IRL. For example, the prevalence of sexism online increases the *acceptability* of gendered harm, with the effects of virtual discrimination transferred to the physical world.[299] It thus influences sexist attitudes and behaviour offline.

Theories on cyber norms aim to explain informal influences on human behaviour in cyberspace and may serve to broaden the discussion on the exacerbation of certain types of injury online. Such theories espouse a social constructivist viewpoint, contending that technology is shaped by human action and vice versa, and thus form part of the cyberfeminist discourse.[300] From this technosocial perspective, gender and power, that is, social norms, both shape and are shaped by technology.[301] It abandons the binary idea of technology as solely a tool, or as the main stimulus of gender-based online harm, considering that, '. . . technology is both a source and a consequence of gender relations'.[302] For example, the advent of social media has brought significant changes in how individuals meet, communicate and interact, that is, social behaviour.[303] The content of communication is also limited by the methods of communication available in society. That is, the ideas that can be communicated vary depending on the means, be it smoke signals, radio, newspapers, mail, or the Internet and, as a consequence, the degree and form of harm.[304] Meanwhile, the manner in which ICTs are utilised in practice also shapes technological development, that is, social norms affect design.

Certain aspects of the Internet are considered particularly important for the development of cyber norms: the anonymity available to Internet users, the continuous flow of information and group mentality in closed fora. These features appear to facilitate and heighten the gravity of offences.[305] Anonymity increases the risk of gender-based offences being committed as it has a disinhibiting effect, instilling the sense that the Internet is disconnected from real life and allowing users to push the social boundaries of acceptable speech.[306] The design of social media encourages more informal speech which, in terms of content, is akin to verbal communication. As argued, the physical distance and the possibility of user anonymity has a

[299] Netanel (2000), p. 457; Gurumurthy (2004), p. 27; Fox et al. (2015), p. 436.

[300] Wajcman (2004), p. 46.

[301] Powell and Henry (2017), p. 23.

[302] Wajcman (2004), p. 7.

[303] UNHRC, '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), para. 12.

[304] Postman (2005), p. 7.

[305] UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. 19.

[306] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 5.

disinhibiting effect, and the speed of communication fosters spontaneous, provocative speech.[307] Poorly formulated thoughts that would not have reached a public forum are made visible to large audiences.[308] Users may thus post statements and content that they would not communicate in other settings.

The digital format also entails that there are fewer social restraints on behaviour. Many of the natural checks of shaming—such as eye contact—are not present on the Internet.[309] Because users who engage in negative behaviour are unable to observe the non-verbal reactions of others, such as facial expressions or other indicators of disapproval, there is less social regulation.[310] Furthermore, the feature of asynchronicity—that feedback is not instantaneous—may entail that sexist messages are perceived as less harmful. Where a negative response is not immediate, the author or the audience may not consider the post as offensive.[311] The common lack of domestic regulation and enforcement as well as the absence of visible authorities monitoring content also exacerbate harmful behaviour.[312]

Furthermore, it is easy to meet likeminded individuals. Groups with homogenous views tend to become more extreme as members gain confidence in their preconceived ideas, entrenching and radicalising views, since groups are more inclined to lack a sense of personal responsibility for their actions.[313] Studies indicate that the digital exchange of information intensifies pre-existing beliefs in a group and the corroboration of a person's ideas engenders confidence, which in turn may lead to extremism.[314] The desire for esteem by peers facilitates the process of norm-building as group norms tend to be stronger than broader social norms.[315] Such mob behaviour is categorised as "deindividuation", meaning that people in groups produce more aggressive, deviant or socially unacceptable behaviour.[316] This is also evident on the Internet, where likeminded individuals may easily convene in groups, noticeable in the substantial growth of extremist groups online.[317] Group dynamics can *inter alia* lead to cyber mobs engaging in harassment, either targeting particular individuals or specific social groups.

Linked to these factors, a generous approach to the freedom of expression is a pervasive cybernorm, which in turn heightens the acceptance of coarser language on the Internet.[318] Verbal abuse is seen as a normal feature of the Internet, different

[307] Lidsky (2018), p. 1903.
[308] Gagliardone et al. for UNESCO (2015), p. 14.
[309] Klonick (2016), p. 1029.
[310] Fox et al. (2015), p. 437.
[311] ibid.
[312] Barak (2005), p. 83.
[313] Keipi et al. (2017), p. 27.
[314] Barak (2005), p. 82.
[315] McAdams (1997), p. 355; Major (2000), p. 76.
[316] Bishop (2013), p. 29.
[317] Silke (2010), p. 33.
[318] Major (2000), p. 105.

from the physical world.[319] Meanwhile, studies indicate that the acceptance of a broad freedom of expression as a cyber norm is gendered.[320] It is more common for women to agree with the statement that people should feel welcome and safe in online spaces, whereas more men agree that it is important for people to be able to speak their mind freely online.[321] In surveys on their attitude towards different types of Internet behaviour, such as "flaming", women valued consideration for others more highly, whereas men assigned a greater value to the freedom of expression and limited censorship.[322] As for appropriate means of addressing online harassment, more men favoured the improvement of policies and tools of online companies, while women were more likely to favour the adoption of more effective laws.[323] This gendered approach to the freedom of expression has an impact on the nature of online interactions, with studies indicating that the rhetorical and linguistic content of male communication more frequently involves an adversarial tone.[324] This aligns with both cultural feminist theory—on the gendered difference in moral reasoning—and radical feminism, where difference arises as a result of the experiences of men and women.

Society is also increasingly digitised, with communication devices frequently at hand. Our identities and social lives are ever more visual and on display, creating a blurring of theoretical divisions of public and private spheres and also the dichotomy of physical and virtual lives.[325] This leads to new cultural norms on privacy, for example, on image-taking and sharing as well as cultural practices of humiliation. [326] This may reduce expectations of privacy and, as a consequence, increase acceptability of intrusions into the private sphere, including in relation to sexual autonomy.

[319] Citron (2014), p. 79.

[320] 61% of women consider that online harassment constitutes a major problem, as opposed to 48% of men, in a research study in the US. Pew Research Center, 'The State of Online Harassment' (2021) <https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/> Accessed 11 March 2022.

[321] ibid.

[322] Herring (1996a), p. 149.

[323] Pew Research Center, 'Online Harassment 2017' (2017). This correlates with theories on women's distinctive characteristics in cultural feminism.

[324] Herring (1996a), p. 146. The reason for differences in communication may be debated, also within feminist legal scholarship. From a cultural feminist perspective, it may be considered a result of biological differences between the sexes, corresponding to studies performed by Carol Gilligan, affirming differences in the moral reasoning of men and women. *See* Gilligan (1982). Her studies indicate that women more frequently adopt a supportive attitude and value interpersonal relationships, whereas men are more inclined toward adversarial communication and individualistic behaviour. In contrast, dominance feminism opposes the approach to sex as determinative in terms of individual attributes, rather focusing on a difference in power structures, expressed through, for example, harassment, sexual violence or bullying, as a result of historical gender hierarchies and the male prerogative. However, both such perspectives affirm gender differences in interactions.

[325] Powell and Henry (2017), p. 100.

[326] ibid., p. 131.

The Internet is changing the *nature* of sexual encounters, impacting on sexual behaviour and thus social norms on sexuality. This is apparent through the practices of sexting, so-called revenge pornography, virtual sexual abuse, and the mainstreaming of hard-core pornography.[327] Women and men increasingly express their sexuality through technological means.[328] With increased sexual interactions and displays of sexuality in the public sphere, the risk of abuse increases. For example, Scotland has in recent years recorded heightened levels of reports of rape, which authorities correlate with the increase in social connections online.[329]

The divulging of personal information has also developed into a social norm as a result of social media.[330] Posted content may be available to the public at large or, by choice, to a more limited group of people. Not only does sharing generate individual psychological satisfaction, it is bolstered by the technological design of the most frequented platforms.[331] This is in part a result of the promotion by social media websites of social interaction, stemming from a business model centred on commodifying personal information.[332] User engagement and traction—meaning attracting new users as well as keeping them interactive on the site—are essential to companies, which encourage users to engage more frequently and share more data.[333] This shifts the boundaries of social norms. User engagement is, for example, promoted by allowing inflammatory and controversial content, such as gender-based hate speech, which prompts users to interact.[334] Social media platforms likewise encourage performative speech, such as exaggeration, by allowing audiences to react, for example, through "likes", comments and shares, which heightens the risk of harmful speech being published.[335] Additionally, the design of social media sites creates a false sense of intimacy, and privacy settings are often complex.[336] Ensuring online privacy, primarily in the form of informational privacy through, for example, the use of an opt-in approach, may thus hamper economic innovation and effectiveness.[337]

---

[327] Murray (2016), p. 423.

[328] Evans et al. (2010), p. 119.

[329] Since more individuals are meeting and initiating relationships on the Internet, for example, through social media and dating apps, where connections progress quickly, many individuals do not take regular precautions when meeting in real life, as the person is not considered to be a stranger. *See* The Scotsman, 'Internet dating blamed for rape increase' (29 May 2017) <https://www.scotsman.com/news/politics/internet-dating-blamed-for-rape-increase-1-4460019> Accessed 11 March 2022.

[330] Rizk (2013), p. 953.

[331] Tene and Polonetsky (2014), p. 79.

[332] Rizk (2013), p. 953; Yanisky-Ravid (2014), p. 74.

[333] Tene and Polonetsky (2014), p. 77.

[334] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 85.

[335] Lidsky (2018), p. 1913.

[336] Marsoof (2011), p. 111; Allen (1999), p. 730.

[337] Rizk (2013), p. 979.

The market also influences the prevalence of harmful material online, not through sanctions of behaviour, but by affecting choices directly when made, such as the price of accessing the Internet.[338] On the Internet, the main commodities are informational products, that is, music, written materials, data and computer programs, with transactions involving information processing. [339] When websites require fees for access or advertisers gravitate towards popular websites, certain types of interactions or information are bolstered. Nevertheless, with most Internet content being free, the constraint of market forces on human behaviour is reduced. The fact that illegal or harmful material is easily accessible and without cost thus affects the propensity of users to download or view such content, be it copyrighted material or pornography.[340] This in turn may affect the character of the material. For example, the possibility of disseminating pornography via the Internet has made sexually explicit material available on an unprecedented scale. Studies indicate that because of the vast market for pornography on the web, competition between sites has made images more violent and degrading.[341] Materials denigrating women that would be found unacceptable in other fora thus flourish online and are moved to the mainstream. Previously few people had access to such extreme images. The mainstreaming of one-dimensional images of women, as hypersexualised and objectified, further enhances existing gender stereotypes.[342] Because of the global nature of the Internet, the impact on norm development is particularly widespread. In turn, it has been argued that pornography has propelled technological development, through investments in certain forms of media.[343]

The existence of cyber norms is an important consideration for legislators in order to ensure that laws accomplish their objectives and are accepted by the community in which they are imposed.[344] The perceived gap between online and offline activity not only causes difficulties in transferring law online, but at times also challenges the legitimacy of law, resulting in non-compliance. Similarly, a failure to regulate the Internet undermines the credibility and effectiveness of laws governing equivalent offline activity, that is, it is difficult to enforce laws/social norms IRL if these are unregulated online.[345] This is, for example, the case concerning sexual violence, sexual harassment and gender stereotyping. However, the degree to which cyber norms should influence the application of human rights law requires careful

[338] Lessig (1999), p. 507.

[339] Elkin-Koren and Salzberger (1999), p. 559.

[340] Spinello (2016), p. 4.

[341] Hughes (2002), p. 129.

[342] UN Women in cooperation with ESCAP, UNDP, UNFPA, UNICEF and WHO, 'Report of the Expert Meeting on Prevention of Violence against Women and Girls' (2012), para. 29.

[343] Johnson (1996).

[344] Major (2000), p. 104. For example, the illegality of downloading music and movies has not been broadly accepted as a norm in cyberspace, generally making such laws inadequate. Arguably, Internet users have opted to make norms, rather than law, their control instrument of choice. *See* also discussion in Ellickson (1998), p. 551; Kohl and Fox (2017), p. 8.

[345] Kohl (2007), p. 8.

consideration in order not to lower the threshold of individual protection, while being sensitive to the context. For example, social norms and the more common use of vulgar speech online have been considered by the ECtHR to reduce the harmful impact of statements, such as defamation.[346]

2.3.2.3 Domestic Law

Law is one of the primary means of constraining human behaviour. However, law—whether domestic or international—and the Internet form an unstable relationship as the development of new technologies is characterised by rapid growth and the legal response is often not up-to-date. The effectiveness of the law is impeded by both Internet architecture and cyber norms, complicating legislative and enforcement efforts. Meanwhile, the perceived or actual lack of regulation exacerbates the prevalence of online harm. If there is a limited threat of sanctions, users are more likely to engage in illegal behaviour, and views and conduct are radicalised. This highlights the necessity of developing context-sensitive legislation, that is, either adapting the interpretation of current laws or adopting new legislation in a manner that ensures its applicability online. At the same time, the law as a blunt tool in regulating online behaviour has been used as an argument for strengthening regulation through code.[347] In response, it is increasingly argued that technological development must be guided by law, including international human rights law, and not the reverse.[348]

Despite widespread ratification of international treaties on women's human rights, lacuna in domestic laws on gender-based offences are common.[349] In relation to *online* harm, gaps arise both as a result of the inapplicability of traditional legal concepts as well as a trivialisation of harm experienced by women, affecting regulation and legal assessments of harm.[350] For example, domestic laws on harassment are frequently limited to the areas of employment and education, excluding the online sphere.[351] Regulation of offences commonly defined as involving physical acts, such as various forms of sexual violence, presents particular challenges, for

[346] *Payam Tamiz v the United Kingdom* App no 3877/14 (ECtHR, 19 September 2017), para. 81; *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* App no 22947/13 (ECtHR, 2 February 2016), para. 77.

[347] Lessig (1997), p. 183.

[348] This is further explored in Sect. 3.4.3.4.

[349] UNHRC, '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), para. 78.

[350] European Institute for Gender Equality (EIGE), 'Cyber Violence against Women and Girls' (2017), p. 3. This is further explored in Chap. 4.

[351] European Commission, European network of legal experts in gender equality and non-discrimination, 'Criminalisation of gender-based violence against women in European States, including ICT-facilitated violence' (2021), p. 10.

example, by not encompassing image-based sexual abuse. Where laws are applicable to online harm, they often disadvantage women by not including the types of harm commonly experienced by women. For example, where domestic laws on hate speech encompass online speech, the definitions generally exclude sexist speech.[352] Furthermore, studies in the Nordic states on online offences indicate that men are more often subjected to threats of physical violence or disparaging remarks on their professional competence, whereas women are subjected to sexual and sexist comments and threats to disseminate intimate images, that is, not threats of violence but of violations of individual integrity and reputation.[353] Meanwhile, domestic laws may not criminalise unfulfilled threats *per se* or often require threats of violence, and are thus more likely to encompass online harm against men.[354] Although states maintain a margin of appreciation in the implementation of international human rights, this area of law may contribute to a harmonisation of standards on Internet regulation, remedying gaps in regulation.

At the same time, it is easier to evade the constraints of law online.[355] While offensive comments to a higher degree are recorded and are thus more easily detected, user anonymity significantly undermines the effective investigation and prosecution of online offences.[356] For example, according to statistics, 96% of crimes committed on the Internet reported to the police in Sweden were not prosecuted, the main issue being user anonymity, with social media companies and ISPs unwilling to disclose data.[357] Meanwhile, surveys indicate that in the majority of cases involving online gender-based violence, the perpetrator is anonymous.[358] Law enforcement at the domestic level also often lacks the technical capacity and tools to effectively investigate online offences.[359] This has gendered

[352] ibid., p. 12.

[353] Bladini (2020), p. 4; NIKK, 'Hat och Hot på Nätet: En Kartläggning av. den Rättsliga Regleringen i Norden Utifrån ett Jämställdhetsperspektiv' (2017), pp. 7–9.

[354] UNHRC, '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), para. 82.

[355] Lessig (1997), p. 183.

[356] This is explored further in Sect. 3.4.2.

[357] Brottsförebyggande rådet, 'Polisanmälda hot och kränkningar mot enskilda personer via internet', Rapport 2015:6 (2015) <https://bra.se/publikationer/arkiv/publikationer/2015-02-02-polisanmalda-hot-och-krankningar-mot-enskilda-personer-via-internet.html> Accessed 11 March 2022, p. 12. The Swedish police held that the main obstacle was the non-release of information by foreign companies, particularly under US jurisdiction.

[358] For example, in an EU survey, 83% of victims of online indecent exposure reported that the perpetrator was unknown. *See* European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 112.

[359] UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. 2; CoE (2018a), 'Mapping study on cyberviolence', Cybercrime Convention Committee, Working Group on cyberbullying and other forms of online violence, especially against women and children (CBG) <https://rm.coe.int/t-cy-mapping-study-on-cyberviolence-final/1680a1307c> Accessed 7 March 2022.

consequences, with a World Wide Web Foundation report indicating that the domestic justice system in 74% of Web Index countries failed to take appropriate action in relation to online gender-based violence.[360] Jurisdictional complexities also undermine the enforcement of domestic laws, in view of the transborder character of online communication. This frequently leads to challenges in applying principles of jurisdiction under international law in the traditional sense, and situations of concurrent jurisdiction of multiple states. Enforcement of domestic laws is furthermore curtailed by wide disparities among states on legitimate restrictions of the freedom of expression. A harmonised approach to illegal content is thus necessary also in this regard.

2.4 Conclusion

In international human rights law, access to the Internet is increasingly approached as an aspect of pre-existing rights, such as the freedom of expression, if not an independent right. This stems from its public sphere attributes and as a platform for ensuring a range of human rights. As a consequence, gender equality on the Internet has mainly been considered from a quantitative viewpoint, focusing on the digital gender gap. Obligations have been placed on states to take steps to develop Internet infrastructure and to enhance physical access and user skills for women. Simultaneously, this rights-based approach to Internet access heightens demands on ensuring substantive equality in terms of online content and conduct. It is also clear that the full range of international human rights applies to the online sphere, including obligations to combat gender-based violence and to eliminate gender stereotyping. Ensuring access for women to the Internet and preventing gender-based online violations are in fact interlinked. The digital divide entails that fewer women access the Internet, which skews the content available, undermining diversity. With more limited contribution by women, gender stereotypes are exacerbated, which in turn fuels gender-based violence. Meanwhile, the prevalence of gender-based violence online impedes equal participation, as women may retreat from this sphere in response to harassment, or become non-participating observers. This has, for example, been a noted reaction of female journalists who are more frequently harassed online than their male counterparts. The broader effects of online gender-based harm—reducing women's participation on the Internet—thus undermines the fulfilment of such human rights as the freedom of expression and the right to employment and education. It also weakens the democratic force of the Internet.

As noted above, empirical studies indicate that gender-based offences on the Internet affect women disproportionately. Whereas early cyberfeminists envisioned the Internet as a space for dismantling patriarchal constraints, studies thus indicate

[360] World Wide Web Foundation, 'Web index report 2014–15', <https://thewebindex.org/report/#chapter_4> Accessed 11 March 2022.

that no disengagement of gender has ensued, as the medium is influenced by gender relations in the non-virtual world, leading to a reproduction of social power hierarchies. These online gender-based offences often consist of the objectification of women and involve constraints on sexual autonomy. In relation to the violations analysed in the book it can be concluded that sexual violence and sexual harassment are considered gender-based *per se*, while defamation and disclosure of personal information may have gendered connotations depending on the form and context. Meanwhile, pornographic material and sexist hate speech are more contested issues in international human rights law but are frequently addressed in relation to the proliferation of gender stereotypes, when depicting women as sex objects or as intellectually inferior. Although these forms of gender-based offences existed prior to the development of the Internet and are thus not a direct result of the medium, the Internet is not a neutral platform. It is increasingly recognised in international human rights law that the Internet not only provides a platform where gender-based violations transpire, but that it also aggravates and facilitates such acts. It is consequently necessary to acknowledge the dynamics of power, domination, and inequality *through* technology.

As argued, regular constraints on human behaviour—such as architecture, norms, the market and the law—may be incapacitated or function differently online. Accordingly, certain technical and design features have gendered objectives or gendered effects, the latter including user anonymity; the asynchronicity, universality and durability of information; the automation of services; and the decentralised structure of the Internet. Meanwhile, processes of developing online social norms and group dynamics generate a sphere where sexist stereotypes and certain forms of gender-based harm are exacerbated. In turn, cyber norms affect gender relations beyond the digital sphere. Since the Internet is not insular from the physical world, the presence of sexism online influences stereotypes also in the physical world. Finally, the lack of domestic and international regulation specific to or applicable to the Internet undermines constraints on user behaviour. The male construction, content, use and regulation of the Internet are thus factors affecting the prevalence of gender-based offences in this forum.

To ensure the effective transposition of provisions on gender-based violence and gender stereotyping to the digital sphere, the relationship between gender and technology thus must be taken into account in domestic and international law, in order to ensure their acceptability and effectiveness. This may affect the interpretation of rights and the content of obligations. In turn, the development of international human rights law provisions in this area should influence Internet design in a manner that reduces gender-based violations, for instance, through means of monitoring and moderating content. The existence of effective regulation may also constrain the development of harmful social norms. There is thus an interplay between the modalities of constraint, with international human rights law fulfilling an important role in setting standards for the values to be implemented.

Nevertheless, as will be explored further in subsequent chapters, a contextual approach in the application of rights has in certain instances involved an undervaluing of the harm of online gender-based violations. This may arise during

the course of rights interpretation, assessments of restrictions of rights or in balancing in conflicts of rights. The attributes of the Internet, particularly involving its public sphere features—as beneficial to democracy and the freedom of expression—have prompted regional human rights law courts and UN treaty bodies to be mindful of restrictions affecting the architecture of the Internet. Meanwhile, the impact of cyber norms on civility have influenced legal assessments of harm in a manner that undervalues its severity. The gendered effects of context-sensitive interpretations of the scope of rights and obligations must thus be considered.

References

Allen A (1999) Coercing privacy. William Mary Law Rev 40:723–757
Arnardóttir OM (2002) Equality and non-discrimination under the European Convention on Human Rights. Brill, The Hague
Bailey J, Telford A (2007) What's so "cyber" about it?: reflections on cyberfeminism's contribution to legal studies. Can J Women Law 19:243–271
Barak A (2005) Sexual harassment on the internet. Soc Sci Comput Rev 23:77–92
Bardall G (2017) The role of information and communication technologies in facilitating and resisting gendered forms of political violence. In: Segrave M, Vitis L (eds) Gender, technology and violence. Routledge, Oxford, pp 100–117
Barendt E (2016) Anonymous speech: literature, law and politics. Hart, Oxford
Barlow JP (2019) A declaration of the independence of cyberspace. Duke Law Technol Rev 18:5–7
Benedek W, Kettemann MC (2013) Freedom of expression and the internet. Council of Europe Publishing, Strasbourg. https://book.coe.int/en/human-rights-and-democracy/5810-freedom-of-expression-and-the-internet.html. Accessed 28 Jan 2022
Berger G (2017) The universal norm of freedom of expression - towards an unfragmented internet. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 27–38
Bishop J (2013) The effect of de-individuation of the internet troller on criminal procedure implementation: an interview with a hater. IJCC 7:28–48
Bladini M (2020) Silenced voices: online violence targeting women as a threat to democracy. Nordic J Law Soc 3:1–42
Bowman CG, Schneider E (1998) Feminist legal theory, feminist lawmaking, and the legal profession. Fordham Law Rev 67:249–271
Bridges A et al (2007) Aggression and sexual behavior in best-selling pornography: a content analysis update. Violence Against Women 16:1065–1085
Charlesworth H (1996) Cries and whispers: responses to feminist scholarship in international law. Nordic J Int Law 65:557–568
Charlesworth H, Chinkin C (1993) The gender of jus cogens. Hum Rights Q 15:63–76
Citron DK (2009) Law's expressive value in combating cyber gender harassment. Mich Law Rev 108:373–415
Citron DK (2010) Civil rights in our information age. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 31–49
Citron DK (2014) Hate crimes in cyberspace. Harvard University Press, Cambridge
Clifford J (2013) Equality. In: Shelton D (ed) The Oxford handbook of international human rights law. Oxford University Press, Oxford, pp 420–445
Conaghan J (1996) Gendered harms and the law of tort: remedying (sexual) harassment. Oxford J Leg Stud 16:407–431

Cook R, Cusack S (2010) Gender stereotyping: transnational legal perspectives. University of Pennsylvania Press, Philadelphia
Crockett M (2016) The internet (never) forgets. SMU Sci Technol Law Rev 19:151–181
Daniels J (2009) Rethinking cyberfeminism(s): race, gender, and embodiment. Women's Stud Q 37:101–124
De Hert P, Kloza D (2012) Internet (access) as a new fundamental right: inflating the current rights framework? Eur J Law Technol 3. https://www.ejlt.org/index.php/ejlt/article/view/123/268. Accessed 28 Jan 2022
Dodge A (2016) Digitizing rape culture: online sexual violence and the power of the digital photograph. Crime Media Cult 12:65–82
Döring N, Mohseni MR (2020) Gendered hate speech in YouTube and YouNow comments: results of two content analyses. Stud Commun Media 9:62–88
Eisenstein Z (1998) Global obscenities: patriarchy, capitalism, and the lure of cyberfantasy. New York University Press, New York
Elkin-Koren N, Salzberger EM (1999) Law and economics in cyberspace. Int Rev Law Econ 19: 553–581
Ellickson R (1998) Law and economics discovers social norms. J Leg Stud 27:537–552
Ertürk Y (2004) Considering the role of men in gender agenda setting: conceptual and policy issues. Fem Rev 78:3–21
Evans A et al (2010) Technologies of sexiness: theorizing women's engagement in the sexualization of culture. Fem Psychol 20:114–131
Fascendini F, Fialová K (2011) Voices from digital spaces: technology related violence against women. Association for Progressive Communications. https://www.apc.org/sites/default/files/APCWNSP_MDG3advocacypaper_full_2011_EN_0.pdf. Accessed 28 Jan 2022
Fidler D (2015) Cyberspace and human rights. In: Tsagourias N, Buchan R (eds) Research handbook on international law and cyberspace. Edward Elgar, Cheltenham, pp 94–117
Fineman M (2005) Feminist legal theory. J Gend Soc Policy Law 13:13–23
Fox J et al (2015) Perpetuating online sexism offline: anonymity, interactivity, and the effects of sexist hashtags on social media. Comput Hum Behav 52:436–442
Franks MA (2012) Sexual harassment 2.0. Md Law Rev 71:655–704
Gagliardone I et al for UNESCO (2015) Countering online hate speech. https://unesdoc.unesco.org/ark:/48223/pf0000233231. Accessed 11 Feb 2022
Gardiner B (2018) "It's a terrible way to go to work": what 70 million readers' comments on the Guardian revealed about hostility to women and minorities online. Fem Media Stud 18:592–608
Gilligan C (1982) In a different voice: psychological theory and women's development. Harvard University Press, Cambridge
Gurumurthy A (2004) Gender and ICTs: overview report. Institute of Development Studies. http://www.bdigital.unal.edu.co/51434/1/1858648408.pdf. Accessed 5 June 2020
Halley J (2006) Split decisions: how and why to take a break from feminism. Princeton University Press, Princeton
Haraway D (1991) Simians, cyborgs and women: the reinvention of nature. Routledge, New York
Henry N et al (2020) Technology-facilitated domestic and sexual violence: a review. Violence Against Women 26:1828–1854
Herring S (1996a) Gender differences in computer-mediated communication: bringing familiar baggage to the new frontier. In: Vitanza V (ed) Cyber-reader. Allyn & Bacon, Boston
Herring S (1996b) Gender and democracy in computer-mediated communication. In: Kling R (ed) Computerization and controversy: value conflicts and social choices, 2nd edn. Academic, San Diego
Herring S et al (1995) This discussion is going too far! Male resistance to female participation on the internet. In: Hall K, Bucholtz M (eds) Gender articulated: language and the socially constructed self. Routledge, New York
Holtmaat R (2004) Towards different law and public policy: the significance of Article 5a CEDAW for the elimination of structural gender discrimination. Ministry of Social Affairs and

Employment, The Hague. https://scholarlypublications.universiteitleiden.nl/handle/1887/41992 . Accessed 28 Jan 2022
Holtmaat R (2012) Article 5 CEDAW. In: Freeman M et al (eds) The UN convention on the elimination of all forms of discrimination against women: a commentary. Oxford University Press, Oxford, pp 141–160
Hughes D (2002) The use of new communications and information technologies for sexual exploitation of women and children. Hastings Women's Law J 13:127–146
Johnson P (1996) Pornography drives technology: why not to censor the internet. Federal Commun Law J 49:217–226
Keipi T et al (2017) Online hate and harmful content: cross-national perspectives. Routledge, Oxon
Kittichaisaree K (2017) Public international law of cyberspace. Springer, Cham
Klonick K (2016) Re-shaming the debate: social norms, shame, and regulation in an internet age. Md Law Rev 75:1029–1065
Kohl U (2007) Jurisdiction and the internet: regulatory competence over online activity. Cambridge University Press, Cambridge
Kohl U, Fox C (2017) Introduction: internet governance and the resilience of the nation state. In: Kohl U (ed) The net and the nation state: multidisciplinary perspectives on internet governance. Cambridge University Press, Cambridge, pp 1–23
Laidlaw E (2015) Regulating speech in cyberspace: gatekeepers, human rights and corporate responsibility. Cambridge University Press, Cambridge
Leiter B (2010) Cleaning cyber-cesspools: Google and free speech. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 155–173
Lessig L (1997) The constitution of code: limitations on choice - based critiques of cyberspace regulation. CommLaw Conspectus 5:181–191
Lessig L (1999) The law of the horse: what cyberlaw might teach. Harv Law Rev 113:501–549
Lessig L (2006) Code and other laws of cyberspace: version 2.0. Basic Books, New York
Levit N, Verchick R (2016) Feminist legal theory: a primer, 2nd edn. New York University Press, New York
Lidsky L (2018) #I U: considering the context of online threats. Calif Law Rev 106:1885–1928
Littleton C (1991) Reconstructing sexual equality. In: Bartlett K, Kennedy R (eds) Feminist legal theory: readings in law and gender. Westview Press, New York, pp 35–56
MacKinnon C (1987) Feminism unmodified: discourses on life and law. Harvard University Press, Cambridge
MacKinnon C (2007) Directions in sexual harassment law. Nova Law Rev 31:225–236
Mahoney K (1994) Canadian approaches to equality rights and gender equity in the courts. In: Cook R (ed) Human rights of women: national and international perspectives. University of Pennsylvania Press, Philadelphia, pp 437–464
Major AM (2000) Norm origin and development in cyberspace: models of Cybernorm evolution. Wash Univ Law Q 78:59–111
Marganski AJ (2018) Feminist theory and technocrime: examining gender violence in contemporary society. In: Steinmetz KF, Nobles MR (eds) Technocrime and criminological theory. Routledge, New York, pp 11–34
Marsoof A (2011) Online social networking and the right to privacy: the conflicting rights of privacy and expression. Int J Law Info Technol 19:110–132
McAdams RH (1997) The origin, development, and regulation of norms. Mich Law Rev 96:338–433
McGlynn C, Rackley E (2017) Image-based sexual abuse. Oxf J Leg Stud 37:534–561
Megarry J (2014) Online incivility or sexual harassment? Conceptualizing women's experiences in the digital age. Women's Stud Int Forum 47:46–55
Meyer C (1999) Women and the internet. Tex J Women Law 8:305–324
Meyerson D (2007) Understanding jurisprudence. Routledge, New York

Miller SF (2003) Prescriptive jurisdiction over internet activity: the need to define and establish the boundaries of cyberliberty. Ind J Glob Leg Stud 10:227–254
Murray A (2016) Information technology law: the law and society, 3rd edn. Oxford University Press, Oxford
Netanel NW (2000) Cyberspace self-governance: a Skeptical view from liberal democratic theory. Calif Law Rev 88:395–498
Nussbaum M (1997) The feminist critique of liberalism. Department of Philosophy, University of Kansas
Nussbaum M (2010) Objectification and internet misogyny. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 68–90
Nyberg AO (2004) Is all speech local? Balancing conflicting free speech principles on the internet. Geo Law J 92:663–688
Palazzini L (2012) Gender in philosophy and law. Springer, Dordrecht
Patton RB (2015) Taking the sting out of revenge porn: using criminal statutes to safeguard sexual autonomy in the digital age. Geo J Gend Law 16:407–443
Plant S (1996) On the matrix: cyberfeminist simulations. In: Shields R (ed) Cultures of internet: virtual spaces, real histories, living bodies. Sage, London, pp 170–183
Post D, Johnson D (1996) Law and borders: the rise of law in cyberspace. Stanford Law Rev 48: 1367–1402
Postman N (2005) Amusing ourselves to death: public discourse in the age of show business. Penguin Books, New York
Powell A, Henry N (2017) Sexual violence in a digital age. Palgrave Macmillan, London
Radacic I (2008a) Feminism and human rights: the inclusive approach to interpreting international human rights law. UCL Jurisprud Rev 14:238–276
Radacic I (2008b) Gender equality jurisprudence of the European Court of Human Rights. EJIL 19: 841–857
Radu R (2019) Negotiating internet governance. Oxford University Press, Oxford
Reed C (2012) Making laws for cyberspace. Oxford University Press, Oxford
Rizk H (2013) Fundamental right or liberty: online privacy's theory for co-existence with social media. Howard Law J 56:951–982
Schauer F (1998) Internet privacy and the public-private distinction. Jurimetrics 38:555–564
Schulte SR (2011) Surfing feminism's online wave. Fem Stud 37:727–744
Schultz T (2008) Carving up the internet: jurisdiction, legal orders, and the private/public international law interface. EJIL 19:799–839
Shepherd T et al (2015) Histories of hating. Social Media + Society July–December 2015, pp 1–10
Shor E, Seida K (2019) "Harder and harder"? Is mainstream pornography becoming increasingly violent and do viewers prefer violent content? J Sex Res 56:16–28
Silke A (2010) The internet & terrorist radicalisation: the psychological dimension. In: Dienel H-L et al (eds) Terrorism and the internet: threats, target groups, deradicalisation strategies. IOS Press, Amsterdam, pp 27–40
Smith G (2017) Cyberborders and the right to travel in cyberspace. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 125–144
Spinello R (2001) Code and moral values in cyberspace. Ethics Inf Technol 3:137–150
Spinello R (2016) Cyberethics: morality and law in cyberspace. Jones & Bartlett Learning, Burlington
Suzor N et al (2017) Non-consensual porn and the responsibilities of online intermediaries. Melb Univ Law Rev 40:1057–1097
Suzor N et al (2019) Human rights by design: the responsibilities of social media platforms to address gender-based violence online. Policy Internet 11:84–103
Tene O, Polonetsky J (2014) A theory of creepy: technology, privacy, and shifting social norms. Yale J Law Technol 16:59–102

Timmer A (2011) Toward an anti-stereotyping approach for the European Court of Human Rights. Hum Rights Law Rev 11:707–738
Tsagourias N (2015) The legal status of cyberspace. In: Tsagourias N, Buchan R (eds) Research handbook on international law and cyberspace. Edward Elgar, Cheltenham, pp 13–29
Tully S (2014) A human right to access the internet - problems and prospects. Hum Rights Law Rev 14:175–195
Turkle S (1995) Life on the screen: identity in the age of the internet. Simon & Schuster, New York
Vera-Gray F et al (2021) Sexual violence as a sexual script in mainstream online pornography. Br J Criminol 61:1243–1260
Vitis L, Segrave M (2017) Introduction. In: Segrave M, Vitis L (eds) Gender, technology and violence. Routledge, London
Wajcman J (2004) Technofeminism. Polity Press, Oxford
Warf B (2017) Alternative geographies of cyberspace. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 147–164
West R (1988) Jurisprudence and gender. Univ Chic Law Rev 55:1–72
Williams W (1991) The equality crisis: some reflections on culture, courts, and feminism. In: Bartlett K, Kennedy R (eds) Feminist legal theory, readings in law and gender. Westview Press, New York, pp 15–34
Yanisky-Ravid S (2014) To read or not to read: privacy within social networks, the entitlement of employees to a virtual private zone, and the balloon theory. Am Univ Law Rev 64:53–104

Chapter 3
Challenges in International Human Rights Law

3.1 Introduction

The previous chapter outlined the general applicability of international human rights law to the Internet and, more specifically, the protection against gender-based offences and the elimination of gender stereotyping. The process in developing the Internet, its architecture, cyber norms and the inapplicability of domestic laws were identified as factors impeding the effective governance of this sphere. Nevertheless, the transposition of international human rights law to the Internet is also hampered by various ideological and practical constraints related to the nature of public international law and the theoretical foundation of international human rights law. Broader issues of relevance to the applicability of international human rights law to interpersonal offences on the Internet will thus be discussed in the following sections with the aim of assessing the neutrality of international law, be it through its creation and/or effects. These fundamental issues to an extent explain difficulties in regulating technology-facilitated harm *per se* and gender-based offences in particular.

The focus of the following parts lies on two aspects: the delineation of what is considered harmful and the regulation of liability on the Internet. *Whether* online behaviour is deemed harmful from a legal perspective and *how* individuals are harmed affect both domestic and international law, that is, whether and to what extent conduct and content is regulated. For instance, it affects the categorisation of harm as a human rights law violation, the balancing in conflicts of rights and the content of state obligations. These sections will accordingly explore theoretical and legal approaches to the concept of harm, how harm is conceptualised from the perspective of the values and scope of rights as well as what balancing exercises and proportionality assessments may indicate in terms of harm, at a general level. Furthermore, the values of rights indicate the boundaries for regulation, to be considered also in conflicts of interests. For example, even if speech or conduct is considered harmful, the values associated with the freedom of expression may override regulation. In delineating the scope of select rights—the freedom of

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7_3

expression and the right to privacy—the online/offline coherence will also be addressed at a general level.

In terms of liability, as the chapter on specific offences in the main details state obligations to regulate individuals, liability will primarily be considered in relation to intermediaries in this section. This explores secondary liability for intermediaries, state obligations to regulate intermediaries and means of controlling content on the Internet.

3.2 What Is Harmful?

3.2.1 Introduction

From a feminist legal perspective, a contextual and socially conscious interpretation of rights is necessary in order to ensure substantive gender equality.[1] Concretely, a gender-sensitive approach includes the recognition of harm of a certain level of gravity, exclusive to or mainly affecting women, as human rights law violations and a contextual evaluation of harm that does not undervalue the experiences of women.[2] The vulnerability of women in accessing human rights is to a degree acknowledged in international human rights law. This has, on the one hand, involved the creation of specific international treaties on women's human rights and, on the other, the interpretation of rights in a gender-sensitive manner. Nevertheless, several of the offences discussed in the book remain contested in some respect, in many cases connected to the question of harm. This includes whether speech in general, or specific forms of speech—for example, sexist speech and pornography—generate individual or social harm, that is, whether such speech is harmful. In cases where harm is recognised, the severity of the harm also engenders disagreement, as does the appropriate balancing is conflicts with the freedom of expression. Additionally, the effect of the online environment on the legal assessment of harm must be addressed, as the Internet affects how individuals interact and what can be communicated. A reconceptualisation of the concept of harm and its assessment may thus be necessary in view of the particular features of the Internet, which is informed by both empirical studies and theoretical frameworks. Arguably, traditional notions of corporeality limited to the physical body is not applicable in a technosocial world.[3] As harm arising from speech is particularly contested, and most prevalent online, this will be the focus of this section.

[1] MacKinnon (2007), pp. 82, 149.

[2] Sjöholm (2017), p. 671.

[3] Powell and Henry (2017), p. 61.

3.2.2 Technosocial Harm

As viewed, the Internet is multifunctional. For example, it may be used as a workspace, for education, as a social sphere, for political engagement and general information-seeking.[4] The consequences of offences occurring in a sphere relevant to such varied facets of a person's life are thus vast, involving both individual and social harm. Online violations are a mixture of the digital and physical worlds and it is clear that the binary of the virtual/real world as a result has begun to erode. Certain offences are physical although the perpetrator does not touch the victim, for example, by coercing a person to remove his/her clothes in front of a web camera or instructing another person to sexually abuse a victim and record it digitally. Technology-facilitated acts can also escalate into physical violence, for example, rape In Real Life (IRL) following grooming in chat rooms. For instance, non-consensual recordings or distributions of private sexual material may be used to pressure and control individuals as, for example, domestic violence or sextortion.[5] It also raises the risk of offline stalking and physical attacks.[6] Psychological harm may in certain instances produce physical consequences, for example, self-harm by the victim, including suicide.[7] Nevertheless, as the Internet is mainly a textual medium, many of the offences are conducted through speech and most commonly generate psychological harm, which affects the categorisation of the injury at the domestic and international levels.[8]

It is clear that not all individuals are equally affected by harmful acts, be it physical or psychological injury.[9] Nevertheless, as will be discussed below, there is arguably greater variations in relation to psychological harm, relative to the individual. Various forms of online offences may have different consequences. However, there are common themes in the effects of online gender-based harm. Several studies indicate that victims of online harassment often experience depression, stress, suicidal thoughts or anxiety, at times with long-term psychological consequences.[10] Reputational harm, through image-based sexual abuse or

[4] Citron (2014), p. 26.

[5] Bjarnadottir (2016), p. 206; Citron and Franks (2014), p. 351; Langlois and Slane (2017), p. 4.

[6] Citron and Franks (2014), p. 350; Suzor et al. (2017), p. 1060.

[7] UNHRC, '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, para. 27.

[8] Fascendini and Fialová (2011), p. 21.

[9] Saha et al. (2019), p. 255.

[10] European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), pp. 89, 114. *See* also an IPSOS Mori poll on online harassment in eight countries, commissioned by Amnesty International in 2017, <https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/> Accessed 8 March 2022. *See* also Nielsen (2002), p. 273; Saha et al. (2019); Citron (2014), p. 10. The UN Special Rapporteur on Violence against Women also notes that common effects include avoidance of the Internet, psychological harm such as depression and fear, physical harm (e.g. when

defamation, may cause psychological injury as victims tend to internalise socially imposed shame.[11] Such offences may also generate harassment and bullying of the victim.[12] Meanwhile, sexual objectification may lead to a self-objectification process for women, with associated psychological harms such as eating disorders and anxiety.[13] Fear is also a common effect, not only that of further verbal harassment, but of an escalation into physical harm, for example, rape or in-person stalking. The fear may lead to a person changing jobs, residence or even name.[14] Reputational harm may also affect the careers of individuals, thus producing economic harm.[15] The online defamation of a person's character may lead to a reluctance of companies to employ the individual in question or to a dismissal. According to a 2009 Microsoft study, 70% of recruiters had rejected candidates based on information found online.[16] Frequently, employers use social networking for recruitment.[17] Hence, a person's Internet presence has an impact on their employment prospects.

For women who engage in public debate on the Internet, the risk of harassment in certain instances leads to a retreat from this sphere or a modification of their behaviour.[18] Common responses include blocking content, adjusting security settings to filter abuse, changing usernames, and diminishing their online presence.[19] For example, reports demonstrate that female politicians and journalists commonly refrain from discussing certain topics for fear of receiving threatening responses.[20] In a survey coordinated by Amnesty International, 76% of women who replied that they had experienced abuse or harassment on a social media platform adapted their

posting the home address of victims) and economic harm. It thus impacts on women's visibility and participation in public life and leads to a democratic deficit. *See* UNHRC, '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), paras. 27, 29. The UN HRC, 'Resolution adopted by the Human Rights Council on 5 July 2018: Accelerating Efforts to Eliminate Violence against Women and Girls: Preventing and Responding to Violence against Women and Girls in Digital Contexts' (17 July 2018) UN Doc. A/HRC/RES/38/5 also gives an overview of the effects on women.

[11] Citron and Franks (2014), p. 364.

[12] Suzor et al. (2017), p. 1060.

[13] Ford et al. (2015), p. 255; Fredrickson and Roberts (1997).

[14] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 7.

[15] Hill (2015), p. 121.

[16] Microsoft Study, 'Online Reputation in a Connected World, Cross-Tab' (2010) <download.microsoft.com/.../DPD_Online%20Reputation%20Research_overview> Accessed 11 March 2022.

[17] Roulin (2014), p. 81.

[18] UNHRC, 'Report of the Working Group on the Issue of Discrimination against Women in Law and in Practice' (19 April 2013) UN Doc. A/HRC/23/50, para. 66.

[19] Vitis and Segrave (2017), p. 8.

[20] UNESCO, 'World Trends in Freedom of Expression and Media Development: 2017/2018 Global Report', p. 157; OSCE, 'New Challenges to Freedom of Expression: Countering Online Abuse of Female Journalists' (2016).

Internet use.[21] Withdrawal from the Internet may have professional costs as it in certain instances results in lost business opportunities, for example, through blogs or networking.[22] At the level of the individual, the consequences of online harm thus involve tangible harm in the form of constraints on individual autonomy—including physical or psychological integrity—social and political participation, and economic security.

These individual experiences aggregate into broader, group-based consequences, negatively affecting gender equality and democracy. Political, social, and economic life is increasingly digitised and ICT-based violence may reduce online involvement in such areas. From a broader perspective individual harm thus affects democratic participation, generating a loss of a diversity of viewpoints and exacerbating social and economic inequality. However, the group-based effects of speech are particularly contested as the empirical verification of a causality between speech and social harm is limited. This includes the impact of pornography, sexist hate speech and gender stereotyping on social norms and culture.[23] Such a link is mainly inferred from theoretical arguments, primarily in relation to gender equality, discussed below.[24] Although empirical studies may not correlate with legal assessments of harm, the latter generalisations are informed by subjective experiences. However, causality—which is not the same as correlation—is complex, given the existence of possible concealed or immeasurable conditions affecting the outcome.[25] It should thus be borne in mind that the causality between certain acts/speech and harm may not be linear and empirically conclusive. Nevertheless, statistics demonstrating that certain violations disproportionately affect a particular social group may certainly infer such social harm as inequality.

Despite these individual and group-based consequences, the online sphere is in several regards considered a factor mitigating harm in domestic and international law. Harm experienced through new technologies is frequently unregulated at the domestic level, in part due to such offences being perceived as less severe.[26] For example, the Group of Experts on Action against Violence against Women and

[21] 32% of women indicated that they refrained from posting on certain issues. *See* IPSOS Mori poll on online harassment in eight countries, commissioned by Amnesty International in 2017 <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022. *See* also BRÅ, 'Hot och kränkningar på internet' (2019), <https://bra.se/download/18.62c6cfa2166eca5d70e5198e/1614334508081/2019_Hot_och_krankningar_pa_internet.pdf> Accessed 11 March 2022, p. 23, in which 20% of victims of online harassment refrained from an online activity.

[22] CoE, 'Seminar Combating Sexist Hate Speech: Report', 10–12 February, EYC, Strasbourg, p. 22.

[23] For example, there are ethical considerations that place certain restrictions on conducting empirical research on the harm of hate speech. *See* Jay (2009), p. 81.

[24] Section 3.2.3.2.2.

[25] McGrogan (2016), p. 631.

[26] UNHRC, '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), para. 68.

Domestic Violence (GREVIO) of the CoE has noted that one of the main impediments in regulating online gender-based harm against women is the failure to recognise its social, economic, psychological and participatory effects.[27] This is linked to the view of the Internet as disembodied. Cyberspace is often understood as being ethereal and separate from everyday life, affirming an online/offline dichotomy.[28] It is in turn reinforced by the cyberlibertarian ideology, which considers that the value of the Internet to liberal democracy rests on it being a separate and unregulated sphere. However, with the world of the Internet viewed as simulated, as opposed to real and material, a dichotomy between online and offline worlds is created also *vis-à-vis* harm. This produces limiting dualisms: online/offline and physical/virtual.

This in part stems from the perceived ability of users to avoid harm online, which reduces the incentive for states to regulate or enforce legislation.[29] Anecdotal and empirical studies indicate that domestic authorities frequently place the responsibility on women to avoid offences in cyberspace. [30] The response by domestic law enforcement often involves providing cyber safety tips, such as changing telephone numbers or computer passwords; adopting androgynous or male avatars; refraining from taking/sharing intimate photographs; deactivating accounts or blocking individuals.[31] It is also not uncommon that law enforcement officials advise victims to stay offline.[32] The solution is thus perceived as women hiding female characteristics or retreating from the Internet, in order to reduce the risk of being victimised.

Furthermore, by approaching digital offences from the standpoint of morality and thus offence, rather than physical or psychological harm, the morality of the victim is often questioned. Victims are in certain instances held responsible for their own abuse, for example, when they blog or write about controversial topics.[33] In situations of "revenge porn", it is frequently implied that the victim is to blame by having consented to the taking of intimate photographs or videos. When photographs are taken without consent, authorities tend to react more strongly.[34] The implication is

[27] CoE, Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', adopted on 20 October 2021, para. 16.

[28] Powell and Henry (2017), p. 50.

[29] Citron (2009b), p. 397.

[30] UNHRC, '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), para. 68; UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. 2. Ambiguous delineations of state jurisdiction and limited intermediary liability are also factors.

[31] Powell and Henry (2017), p. 244.

[32] Citron (2014), p. 84.

[33] UNHRC, '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), para. 68; Citron (2014), p. 19.

[34] Hill (2015), p. 123.

that the person, through his or her actions, has reduced reasonable expectations of privacy and the harm suffered is thus diminished.[35] These gendered precepts—viewing violations against women as less harmful since they can be avoided or mitigated, in addition to a narrow view on the appropriate boundaries of female sexual autonomy—are part of the historical trivialisation of gender-based offences at the domestic level. As noted by Danielle Keats Citron, it stems from a reluctance to acknowledge harm where women arguably could have mitigated injury, for example, in cases of domestic violence, sexual harassment or rape.[36]

Additionally, the use of digital communication is at times a factor used by domestic authorities to trivialise the harm experienced in interpersonal offences, with the argument that the perpetrator may be a stranger in a distant location, implying that the fear is less well-founded.[37] The medium accordingly reduces the harmful effect of speech. However, fear may also arise from not knowing the identity of the perpetrator.[38] Additionally, there are indications that the Internet in certain instances may enhance the level of psychological harm. Misogynist remarks, derogatory nicknames and comments about appearance were, according to an empirical study on sexually harassing behaviour, considered more harassing online than in IRL by victims.[39] The opposite was the case in instances of solicitation of sexual acts. Regarding the first mentioned category, such behaviour was deemed more threatening when expressed in writing. Writing a comment may indicate that the content is intentional and serious, with little context in which to interpret its meaning. [40] In relation to the solicitation of sexual acts, online requests were deemed easier to avoid and less threatening, whereas IRL the victim is in the presence of the harasser and is thus a captive audience.[41]

Furthermore, as argued by the ECtHR, '. . .the information emerging [from the Internet] . . .does not have the same synchronicity or impact as broadcasted information'.[42] Access to information online is mainly a proactive process by individuals, and thus, potentially, not as intrusive. Individuals exercise more control over the information they receive than through traditional media, given that they can search

[35] Citron (2014), p. 77.

[36] Citron (2009b), pp. 394, 400.

[37] For example, it is argued that demeaning words on the Internet should not be treated as severely as those uttered IRL, bearing in mind the distance between the perpetrator and potential victim; the fact that the victim often can choose which websites to visit and there is thus an increased possibility of avoiding such speech and a lower risk of threats being carried out. *See* Brennan (2009), p. 127. For example, Barak argues that the '. . .virtual environment enables people to provide themselves with relative protection from. . . [sexual harassment] and other aggressions'. *See* Barak (2005), p. 83.

[38] Powell and Henry (2017), p. 10; UNHRC, '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), para. 29.

[39] Biber et al. (2002), p. 33.

[40] ibid., p. 38.

[41] ibid., p. 38.

[42] *Animal Defenders International v the United Kingdom* (2013) 57 EHRR 21, para. 119.

for particular content; search engines and other platforms may adapt to personal preferences; and users can install filters. A person as a result has greater leeway, on the basis of personal morality, to avoid certain content, a factor diminishing harm.[43] This reduces the justification for restrictions, particularly concerning harmful but not illicit material, such as obscene pornography.[44] At the same time, information not suitable for children is easily accessible with few restrictions, and certain material, such as spam and harassment, is not the result of proactive choices. It must also be considered that certain material is detrimental to gender equality, such as harmful pornography, and thus beyond the scope of personal morality and choice.

3.2.3 Theories on Harm

3.2.3.1 The Concept of Harm and the Limits of Intervention

Although in part external to international human rights law, and with varying approaches in domestic legal systems, certain theories on harm have nevertheless been influential in the interpretation of rights by international human rights law bodies.[45] The legal assessment of harm is linked to theories on the proper limits of governmental power, indicating which protective interests and injuries warrant state intervention. Whereas the discourse on the public/private dichotomy considers such a division from a general standpoint *vis-à-vis* subject matter, the approach to the harm principle considers state intervention in relation to the nature and degree of harm.

Criminal law is in numerous liberal democracies influenced by the Anglo-American legal tradition, premised on John Stuart Mill's harm principle that an act that does not cause harm to others should not be prohibited.[46] For example, according to Mill, a person should have the freedom to speak and act in accordance with his/her wishes unless it causes harm, the aim being to restrict state intrusion on individual liberty.[47] This emanates from such key concepts of liberalism as rationality, individualism, and freedom from interference by the state, as well as protection of the private sphere.[48] Consideration of collective consequences is thus limited.

---

[43] Berger (2017), p. 36.

[44] ibid., p. 35.

[45] *See*, for example, an analysis of the harm principle by Mill in relation to the concept of hate speech in international human rights law in Pejchal (2020).

[46] 'That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others'. *See* Mill (1859), p. 22. *See* also Feinberg (1984), p. 31. According to Hyman Gross, 'it is harms that make conduct criminal, because the conduct produces or threatens the harm, or even in some cases constitutes the harm'. *See* Gross (1979), p. 114.

[47] Mill (1859), p. 101.

[48] Kofman (2003), p. 395.

As international human rights law is greatly influenced by this ideology, it also reflects such characteristics.[49] The liberalist approach to the harm principle does as such not allow for regulation on the basis of moralism or paternalism, for example, in order to prevent individuals from harming themselves.[50] This posits that harm can be defined in a non-moralistic way. Furthermore, harm is primarily understood as involving physical injury or harm to property.[51] Moreover, harm may not be merely speculative, although the principle also includes certain forms of harm that have not yet materialised, that is, significant risks of harm.[52] In terms of causality between the conduct and harm, *conditio sine qua non* is a common principle in establishing factual causation in both Anglo-American law and Continental legal systems, requiring a necessary condition or cause.[53] Again, this is mainly applied in relation to individual harm.

There are different approaches to what constitutes "harm" and it is accordingly an ambivalent concept. Nevertheless, it is clear that harm is not synonymous with "hurt".[54] Not everything that "hurts", whether physically or psychologically, is considered harmful.[55] Joel Feinberg, whose theories are widely employed in legal theory, defines harm abstractedly as a 'wrongful set-back of interests'.[56] These interests include "welfare", which protects, for instance, emotional stability, a tolerable social environment and a certain amount of freedom from coercion.[57] Harm is accordingly distinguished from taking offence—an unpleasant state of mind, such as anger, irritation or disgust caused by the behaviour of another—and is mainly considered subjective.[58] Experiences of temporary discomfort or pain are thus not included.[59] Nevertheless, the assessment of harm may be both subjective, such as in tort, or generalised, as a basis for adopting certain criminal laws. As such, a generalisation of human experiences and reactions in a certain context must be made. For example, while the question of how women experience sexual violence is empirical, the delineation of what constitutes the harm of rape is theoretical.[60] Harm and experience are thus not inextricably linked. However, acts are perceived as

[49] *See*, for example, in relation to the ECHR: Joppke (2013), p. 102; Nussberger (2020), p. 185.

[50] Maris (2013), p. 4. It should also be noted that feminist theorists have opposed regulation solely on the basis of morality, as this may be used to restrict women's rights. Thus, where states prohibit obscene pornography due to immorality, the foundation rather than the restriction is contested from a feminist viewpoint.

[51] McCloskey (1998), p. 54.

[52] Turner (2014), p. 305; Feinberg (1984), p. 11.

[53] Persak (2007), p. 43.

[54] Feinberg (1984), p. 46.

[55] For example, injuring yourself while exercising may hurt but it is not harm in the legal sense.

[56] Feinberg (1984), pp. 35–36.

[57] ibid., p. 37. *See* also Gross (1979), p. 89.

[58] Feinberg (1985), p. 5. According to Martha Nussbaum, disgust is an inadequate basis for law-making. See Nussbaum (2004), p. 14.

[59] Bell (2021), p. 165.

[60] Baber (1987), p. 125.

violations precisely because they *typically* give rise to harm, even when they do not in a particular case. The above mentioned empirical studies on how online acts/speech are experienced by victims are thus useful in influencing the legal concept of harm but other aspects, such as the perceived severity of the act as well as the balancing of competing interests, ultimately determines its scope.

Accordingly, both speech and physical conduct may in theory generate harm if entailing a set-back to interests. However, whether the harm rationale encompasses, for example, psychological harm, is a question that divides legal theorists and philosophers—with the dominant discourse on harm minimising non-physical injury—as does the requisite level of harm to warrant regulation. This is particularly contested in relation to the harm of speech, which is mainly construed as emotional.

3.2.3.2 The Harm of Speech

3.2.3.2.1 Harm Relative to the Values of Speech

Theorising the harm of speech is challenging, both in relation to individual and social harm. From the perspective of the freedom of expression, there are two broad views against restricting speech, correlated to the concept of harm. One is the minimalist view that speech causes no harm that cannot be redressed by more speech, or causes less harm than conduct. As liberal democracy is premised on state constraint unless harm is caused, the view that speech is harmless would exclude its regulation *per se* and thus many of the offences prevalent on the Internet, which are most often speech-based. Meanwhile, the maximalist view acknowledges that speech can be harmful but that this is trumped by the overriding value of speech.[61] Accordingly, from these viewpoints, speech is either considered costless or, alternatively, priceless.

The view that speech is less severe than physical conduct can either be a standalone argument or addressed in conjunction with theories on the freedom of expression. For example, it is an implied aspect of the autonomy-based value of this right, since it is assumed that autonomy is curtailed in instances of physical injury but generally not by speech.[62] The distinction between speech-based and physical harm tends to focus on difference in duration and intensity. However, this has been disputed, bearing in mind that psychological effects, whether arising from speech or conduct, may linger for longer periods of time, for example, a common consequence of rape.[63] Moreover, the intensity cannot be categorised on the basis of its physical or psychological effects since the former may be minor and the latter particularly grave, depending on the circumstances. Harmful speech may also cause physical

[61] Brison (1998), p. 40.

[62] Schauer (1993), pp. 640–641. It is also implied in the marketplace of ideas theory developed by Mill.

[63] Brison (1998), p. 44.

hurt, for example, as a result of anger and humiliation.[64] Additionally, in response to the notion that the harm of speech can be redressed with more speech, it has been argued that there is insufficient time between speech and the resulting injury for the remedy of more speech to prevent the harm, and this would thus not eliminate the initial injury.[65]

Furthermore, whereas the effects of certain types of physical conduct are easily generalised, it is argued that psychological injury encompasses a broader spectrum of individual reactions than physical harm, dependent on the beliefs and circumstances of the particular listener.[66] For example, certain research in psychology disputes the suitability of universally restricting speech, as harm from offensive speech is contextually determined.[67] Complex variables affect the impact of speech, such as the relationship between speaker and listener, the vulnerability of the person, the words used, and body language etc.[68] Accordingly, the effects of speech are relative to the individual listener and his/her culture.[69] It is thus considered more difficult to ascertain and measure such hurt.[70] This is premised on the approach that individuals are more in control over their mental than physical reactions, that is, it is the responsibility of the listener as opposed to the speaker.[71] Arguably, the harm of sexual harassment is particularly subjective, resulting in the intangible harm of mental distress.[72] In many common law states there has thus historically been a reluctance to adopt tort provisions for the intentional infliction of emotional distress in the absence of physical harm as it has been considered difficult to prove psychological injury, with a perceived risk of fictitious claims and widespread litigation.[73] Accordingly, '. . .harm caused by words is nebulous, easily exaggerated, and readily contrived'.[74] As argued by Frederick Schauer, this stems from the standpoint that it is more reasonable to expect a person to alter his/her beliefs and avoid distress than to guard against physical injury, and is thus not value-neutral.[75] However, conclusions on hurt are less relevant than assessments of harm, which may be equally difficult to ascertain in relation to physical injury. It should be acknowledged that both speech and physical injuries are context-dependent and open to interpretation. For example, several offences are distinguished primarily on the basis of non-consent, such as sexual violence, requiring contextual assessments of the mindset of both victim and

[64] ibid., p. 50.
[65] ibid., p. 43.
[66] Dworkin (1977), pp. 200–203; Rutzick (1974), p. 7.
[67] Jay (2009), p. 81.
[68] Jay and Janschewitz (2008); Jay (2009), p. 81.
[69] Butler (1997), p. 13.
[70] Rutzick (1974), p. 7.
[71] Brison (1998), p. 53.
[72] Conaghan (1996), p. 428.
[73] Stone (2010), p. 189.
[74] Bennett (2016), p. 498.
[75] Schauer (1993), p. 651.

perpetrator.[76] Harm may thus be connected to how speech/acts are experienced both in relation to speech and physical acts.

Even if certain speech is recognised as harmful, the freedom of expression may override such harm, mainly due to the approach that the injury is less harmful than restricting this right, embodied by the maximalist viewpoint. This approach considers the freedom of expression to be a particularly valuable right. There are multiple reasons, either explicit or that can be deduced, for its prominent position, primarily connected to rights theories on the values of speech. As will be discussed in Sect. 3.3.4, this involves, for example, the integral connection between speech, democracy, and individual autonomy. The maximalist approach is not explicitly reflected in international human rights law, in which the freedom of expression is a qualified right subject to the same state restrictions as limited rights in general. However, special protection can be implied on the basis of how conflicts of rights tend to be resolved by international human rights bodies.[77] Balancing exercises indicate whether harm is recognised from a legal standpoint as well as the importance attached to such harm, that is, whether it is considered minor in relation to competing interests. For example, the lack of recognition of individual or social harm on the Internet partly arises from considering the protection of the freedom of expression supreme. This is linked to the presumed social benefits of the Internet, such as its virtues to democracy, generating a protective stance of this forum.

3.2.3.2.2 Harm to Gender Equality

Addressing the group-based effects of speech is particularly relevant in relation to gender-based harm. Such a connection is increasingly considered in international human rights law with, for example, the CoE noting the '. . .fundamental role of language in forming an individual, and the interaction which exists between language and social attitudes', in relation to sexism.[78] Accordingly, '[t]he sphere of language has become. . .a domain in which to interrogate the cause and effects of social injury'.[79] However, this view is contested and often not theorised.

Harm to gender equality can persuasively be analysed from two theoretical standpoints: feminist legal theory and linguistic philosophy. Feminist legal theory

[76] *See* argument in Brison (1998), p. 55.

[77] *See* Sect. 3.3.4.

[78] CoE, 'Recommendation No. R (90) 4 on the Elimination of Sexism from Language' (Adopted by the Committee of Ministers on 21 February 1990 at the 434th meeting of the Ministers' Deputies). The Committee of Ministers of the CoE requests member states to promote the use of language reflecting the principle of equality and to take measures with a view to: '1. encouraging the use, as far as possible, of non-sexist language to take account of the presence, status and role of women in society, as current linguistic practice does for men; 2. bringing the terminology used in legal drafting, public administration and education into line with the principle of sex equality; 3. encouraging the use of non-sexist language in the media.'

[79] Butler (1997), p. 71.

considers that the costs of non-restraint of offensive speech are uneven, gendered and frequently born by vulnerable groups.[80] As mentioned above, the traditional theoretical perspectives on the value of the freedom of expression and the legal concept of harm are premised on liberalism. First, liberal theory mainly recognises violations against individuals, entailing that group-based harm is not sufficiently acknowledged, although certain treaties have been developed for the protection of vulnerable groups.[81] The formal equality approach associated with liberalism limits the recognition of certain offences, some of which are exclusive to women and others predominantly experienced by women. Similarly, the lack of contextualisation of harm leaves the experiences of certain social groups outside the legal realm. The central focus on autonomy presumes that all individuals experience similar degrees of independence. Vulnerability emanating from the economic, political and social status of certain groups is rarely recognised by, for example, regional human rights law courts, leading to limited acknowledgment of discriminatory laws and practices and, again, structural constraints on autonomy. For example, liberal theory fails to identify the harm of pornography because its conception of injury is 'individuated, atomistic, linear, exclusive, isolated, narrowly tortlike – in a word, positivistic' and thus fails to see how individual incidents manifest social structures of inequality.[82]

Furthermore, the types of speech invoking claims of gender inequality—mainly involving pornography, sexism and harassment—are often considered to merely cause offence. According to feminist theories, the harm principle concretises a conception of the good life which is as non-neutral as the content of international human rights law, with the liberal tradition engaged in 'denoting an injury, typically inflicted upon the body, which can be identified independently of both the context in which it takes place and the understanding of the experience from the point of view of the people involved'.[83] Recognition is further constrained by approaches to causality. For example, even when separate incidents do not rise to the requisite level of harm, such as merely causing insult, speech may generate accumulative harm warranting state intervention.[84] This can be understood as harm done by a group, that is, not specifically to a group, where an isolated incident may not reach the requisite threshold but produces a general harm.[85] Pornography is an example of this, where the solitary consumption by one individual may not be harmful but, arguably, a pervasive culture of pornography is.[86] The traditional causal view

[80] *See*, for example, Meyers (1995); Matsuda (1989), fn. 275.

[81] For example, children, refugees, women and indigenous peoples. At the same time, liberalism is not a single position evident, for example, in the ideas espoused by Kant, Mill and Rawls, with certain theories more accommodating of female harm. *See* Nussbaum (1997), p. 3.

[82] MacKinnon (1989), p. 208.

[83] Munro (2007), pp. 12–13.

[84] Kernohan (1993), p. 51.

[85] ibid., p. 52.

[86] ibid., p. 53.

embodied by, for example, Feinberg's concept of harm, tends to focus on individual injury rather than on how speech or acts contribute to the formation of harmful norms. The link between gender stereotypes and gender-based violence would as such not be recognised. Additionally, the harm principle generally requires that material is necessary and sufficient for the harm to ensue. A multiple causation model where harm occurs as a result of accumulated events is more realistic.[87]

The liberalist approach to individualism, autonomy, consent and equality thus produces limitations on the recognition of various forms of harm associated with women, with particularly disadvantageous effects on the Internet. Accordingly, feminist legal scholars call for a contextual and more nuanced approach to harm that is not premised on the individual as naturally unencumbered but acknowledges gender as a factor and considers injury beyond physical hurt. The balancing between abstract principles may result in gendered pre-disposed solutions, such as in favour of the freedom of expression. The medium of communication should, from such a viewpoint, also be considered in the assessment of harm. Nevertheless, feminist theories addressing harm from the perspective of gender equality are multiple and, in certain instances, conflict.[88] This is also the case in relation to the harm of speech, such as pornography, discussed further in Sect. 4.5. The means of addressing harm from a gender perspective may thus vary.

Meanwhile, theories in linguistic philosophy—more specifically on speech acts—serve to categorise different types of speech along a spectrum of speech and conduct, on the basis of their nature or effects on individuals, social groups or society, and thus the causality to harm.[89] This includes the categorisation of certain forms of speech as discrimination. According to speech act theories, a distinction is made between acts performed through words and acts performed as a consequence of words. The same words may have different force depending on the context. Most relevantly, *perlocutionary* communication is speech that leads to injurious consequences but is not itself synonymous with the effect. It presumes causality between speech and its effect, which may be difficult to establish empirically.[90] It includes such effects as convincing, persuading or deterring another person.[91] The effect may vary depending on the listener. Meanwhile, *illocutionary* speech is synonymous with the immediate effect, that is, it may constitute the harm itself.[92] Subordination and silencing may be the objective of illocutions.[93] The categorisation of speech as illocutionary may thus support the prohibition of such speech as it constitutes an act

[87] Russell (1998). *See* also Mikkola (2019), p. 42.

[88] *See* Sect. 2.2.2.1 on the concept of equality.

[89] The theories were in the main developed by J. L. Austin and J. Searle. *See* Austin (1975); Searle (1969). *See* also description in McGowan (2003), p. 155.

[90] McGowan (2003), p. 155. Meanwhile, locutions are considered statements that assert facts or values.

[91] Austin (1975), p. 109.

[92] Butler (1997), p. 39. This requires that the person speaking is in a position of authority.

[93] Austin (1975), p. 181.

already prohibited under law, such as discrimination.[94] To a degree, the theories correlate with norms on direct *versus* indirect discrimination.[95]

These theories may be applied in the analysis of multiple forms of gendered harm on the Internet. For example, according to the European Court of Human Rights (ECtHR), pornography is a form of speech and thus enjoys protection as such, although states are allowed to regulate obscenity.[96] In contrast, certain scholars categorise pornography as speech acts, constituting either perlocutionary or illocutionary speech. Accordingly, the perceived perlocutionary effect is that viewers of pornography are more inclined to consider women subordinate and more likely to accept rape myths and perpetrate sexual violence.[97] As illocutionary speech, pornography is viewed as an act of gender discrimination.[98] The production of pornography reinforces women's pre-existing inferior social status.[99] It may render certain forms of speech unspeakable, such as leading certain women to feel as if they cannot refuse sex. Likewise, refusal may not be perceived as genuine.[100] However, this approach can be modified to pertain to certain forms of pornography.[101] It should be noted that such viewpoints are contested, as there is no consensus on the illocutionary force of pornography, and its effect may in practice require a demonstration of harm.[102] This categorisation also extends to image-based sexual abuse, particularly involving the distribution of sexual assault videos/images. Such reinforce existing harmful norms on sexuality, which allows for sexual violence to be excused and viewed as inevitable.[103]

Similarly, hate speech may constitute either perlocutionary or illocutionary speech.[104] Hate speech as perlocutionary considers its effect in depriving individuals of rights and liberties, such as access to education and employment. In terms of hate

[94] MacKinnon (1993), p. 30.

[95] In using the example of women's economic disadvantage, this may occur through illocutionary speech, for example, domestic law preventing women from owning property. It may also arise as a consequence of biased hiring practices, giving preference to male applicants, that is, through perlocutionary speech. *See* Langton (2011), p. 429.

[96] *See*, for example, *Pryanishnikov v Russia* App no 25047/05 (ECtHR, 10 September 2019).

[97] Langton (1993), p. 306; McGowan (2003), p. 182.

[98] Butler (1997), p. 72. For example, Catharine MacKinnon considers that pornography subordinates and silences women. However, her conceptual approach has moved from a perlocutionary to an illocutionary model throughout her scholarly career. *See*, for example, MacKinnon (1987), p. 194; MacKinnon (1993), p. 30. Such speech may also have an unintended illocutionary force, as a form of unconscious conditioning. *See* McGowan (2003), p. 180.

[99] Butler (1997), p. 18.

[100] Langton (1993), p. 320.

[101] Saul (2006), p. 239.

[102] Stark (1997), p. 283. It has also been criticised from a speech-act viewpoint, in that there is no indication of specific intent nor authority of producers to subjugate women, thus removing the possibility of its exertive effect. *See* McGowan (2003).

[103] Dodge (2016), p. 67; Powell and Henry (2017), p. 106.

[104] Butler (1997), p. 18.

speech as illocutionary, it does not solely *reflect* social domination, but *enacts* subordination by placing the subject in an inferior position.[105] This depends on pre-existing social positions of dominance and subordination.[106] For example, the Rabat Plan of Action on Hate Speech qualifies it as speech acts.[107] However, as noted by Judith Butler, this requires distinguishing between injuries that are '. . .socially contingent and avoidable, and kinds of subordination that are. . .the constitutive condition of the subject', again challenging the causality between speech and its effects.[108]

Additionally, sexual harassment—including sexually derogatory comments, propositions and jokes—defines an individual as inferior and affects his/her self-identity.[109] Certain empirical psychology studies indicate that how harm is experienced in relation to harassment is gender-related and encompass perlocutionary or illocutionary effects.[110] In a study where female and male participants were exposed to either neutral or sexist comedy skits—portraying women as sex objects or in other traditional gender roles—women who viewed the latter expressed a greater level of self-objectification, thus affecting their mental health.[111] In contrast, men were largely unaffected by the content. Women may thus find sexually explicit jokes and photos more harmful than men, which may be due to social norms and previous life experiences.[112] For example, because of a higher risk generally of being subjected to rape, threats of rape may be perceived as more realistic for women.[113] Additionally, the theory of ambient sexism considers environments in which sexism is prevalent as harmful regardless of whether an individual is directly targeted, through the exacerbation of sexist attitudes IRL.[114] Studies in psychology affirm that prior exposure to sexist jokes leads to a greater tolerance of sexist behaviour.[115]

Although speech act theories were developed with respect to spoken communication, they have also been applied to the context of the Internet, such as in relation

[105] Matsuda (1989), p. 2358. Judge Francoise Tulkens of the ECtHR has also argued that hate speech is not only an opinion but an act. *See* Tulkens (2012), p. 281.

[106] Matsuda (1989), p. 2332.

[107] UN HRC, 'Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence', Appendix (Rabat Plan of Action) (11 January 2013) UN Doc. A/HRC/22/17/Add.4, para. 29.

[108] Butler (1997), p. 26.

[109] Fentonmiller (1994), pp. 579, 599.

[110] Fentonmiller (1994), p. 599. It should be noted that such studies on sexual harassment/online harassment are difficult to compare as these offences are not coherently defined, domestically or internationally, and may be based on different types of behaviour.

[111] Ford et al. (2015), p. 260.

[112] Biber et al. (2002), p. 38. This correlates with the cultural feminist viewpoint that when women are subjected to similar offences as men they may, in certain instances, be experienced differently, due to distinct female characteristics. *See*, for example, Gilligan (1982).

[113] Fentonmiller (1994), p. 568.

[114] Fox et al. (2015), p. 436.

[115] Woodzicka and Ford (2010), p. 185.

to Facebook communication and memes.[116] In fact, given the increased use of technology, 'more actions [are] achievable through "mere" words'.[117] The categorisation of all forms of activities online as involving speech, and thus presumptively protected by the freedom of expression, has in fact been criticised.[118] Nevertheless, it should be noted that Internet architecture has an impact on the illocutionary or perlocutionary force of communication, such as the conditional requirements of authority, speaker intent and the effect on the listener.[119]

The categorisation of speech as acts influences the approach to regulation. The liberalist view of John Stuart Mill considers that the most effective way to eliminate subordinating speech is through more speech.[120] State intervention entails that individual agency is denied and assumed by the state.[121] However, from the standpoint of feminist and linguistic theories, when individuals and groups are subordinated, more speech is a futile solution.[122] Subordination is linked to silencing, where marginalised individuals are proscribed from positions of authority, limiting their ability to speak with force.[123] As will be discussed in relation to theories on the freedom of expression, the liberal perspective of individuals as imbued with equal levels of power and opportunities to participate is thus not reflective of reality.[124] State intervention is consequently necessary in order for their voices to be perceived as authoritative.[125]

The categorical approach that certain types of speech have detrimental social effects *per se* is also opposed on several grounds. Arguably, speech is regarded in an inflated manner, no longer as representations of power but as the *modus vivendi* of power.[126] As such, the profound institutional structures of racism and sexism are reduced to the harm of language.[127] Similarly, Judith Butler argues that the effects of speech are context-specific: the meaning changes over time and place, and may have illocutionary force, yet resistance by the hearer may reclaim the impact of speech.[128] One cannot in advance know the meaning a listener will attach to certain words.[129] A

[116] Grundlingh (2018) and Carr et al. (2012).

[117] Citron (2009a), p. 99.

[118] Citron and Franks (2020), p. 56.

[119] Grundlingh (2018). This is further addressed in Sect. 3.3.2.3.

[120] Jacobson (2000), p. 279.

[121] Butler (1997), p. 43.

[122] Langton (1993), p. 325; Butler (1997), p. 39; Asquith (2007), p. 181.

[123] Asquith (2007), p. 182.

[124] As argued by Nicole Asquith, truth is a contingent object constructed out of social positions of power. It requires state intervention to guarantee equal participation. *See* Asquith (2007), p. 186.

[125] Asquith (2007), p. 184.

[126] ibid., p. 74.

[127] ibid., p. 80.

[128] ibid., p. 13. Similarly, Jerome Neu argues that the meaning of words depends on time, geographical space, the identity of the speaker and listener as well as the context. *See* Neu (2008).

[129] Butler (1997), p. 87; Saul (2006), p. 235.

shift in context may exacerbate or minimise the offensiveness.[130] By fixing the meaning of certain statements, for example, through prohibitions in law, the linguistic agency of the listener is removed. This aligns with the approach of psychological injury as subjective. Such a perspective is also evident in international human rights law, where hate speech is viewed contextually, taking into account the impact of statements on a particular audience.[131] A similar argument has been made in relation to pornography, as diverse audiences may interpret the content in different ways.[132] This does not entail that a contextual approach to the categorisation of speech as discrimination is inimical to feminist objectives. As seen above, linguistic theories consider speech in light of existing structural subordination, not as the solitary means of producing discrimination. The historical, social and institutional context of speech is relevant for the analysis of the effects of speech also from a gender-sensitive viewpoint.

3.2.3.3 International Human Rights Law

Although the harm principle is mainly employed as a concept in domestic law, the regime of international human rights law is also premised on protecting individuals against harm that reaches a certain threshold of severity.[133] Rights, similar to Feinberg's concept of "interests", have developed on the presumption that people suffer a particular form of harm if violated. In this sense, the harm is generalised. However, in several regards, whether certain forms of speech or acts are deemed harmful requires a contextual approach also in international human rights law. The assessment of whether speech is categorised as hate speech, defamation or harassment considers such factors as the speaker, audience, content and form. Similarly, the review of the discriminatory effects of such forms of speech is subject to a contextual evaluation.[134] Meanwhile, the evaluation of harm in international human rights law, in the sense of violations of rights, is broader than the liberalist approach described. It encompasses a variety of injuries, including harm resulting from speech to such interests as morals and non-discrimination. For example, in several cases involving pornography, the ECtHR has discussed it in terms of harm to morals, proposing that harm may be caused at any time when a person is confronted with the

[130] Butler (1997), p. 13.

[131] *See*, for example, *Perinçek v Switzerland* App no 27510/08 (ECtHR, 15 October 2015), para. 206.

[132] Saul (2006), p. 238.

[133] In addition, in order to fulfil the admissibility requirements of such institutions as the ECtHR, the individual has to prove "victimhood", that is, that the person is directly affected by an alleged violation. *See* Art. 34 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), 4 November 1950, 213 UNTS 222, entered into force 3 September 1953. The severity of harm is also relevant in relation to the content of reparations.

[134] *See*, for example, *Case Egyptian Initiative for Personal Rights and Interights v Egypt*, ACmHPR, Communication No. 323/06 (1 March 2011).

material.[135] At the same time, the Court is increasingly restrictive in accepting morality as an aim in limiting the freedom of expression.[136] The approach to assessing the severity and causality of the harm of speech in many instances thus correspond with the theories discussed, which impede the recognition of online offences.

International human rights law does not embrace the minimalist approach to the harm of speech, as it recognises certain forms of speech as harmful.[137] However, even with the view that speech may generate harm in the legal sense, certain aspects are contentious. This pertains to, for example, *whether* certain types of speech are considered harmful—such as pornography—linked to the issue of causality, and the perceived *severity* of the harm, for example, the harm of sexist hate speech being less severe than restrictions on the freedom of expression. The form and degree of harm—physical, psychological or economic—affects the categorisation of the violation of a right, such as whether it contravenes the right to privacy or the prohibition on torture. Whereas all human rights are considered of equal importance, the absolute nature and *ius cogens* character of certain norms entail that they in effect produce more extensive obligations for states. The assessment of harm is thus of importance for the determination of whether it falls within the scope of a qualified or an absolute right. The level of harm also has an impact on the potential balancing of interests in conflicts of rights or restrictions of rights. The assessment of whether an interference is legitimate and necessary implies an evaluation of harm. For example, the exposure to certain material may be perceived as causing minor harm, outweighed by the benefits of an uncensored Internet.[138] A person may thus be injured, even within the scope of a right, without there being a human rights law violation.[139]

In international human rights law, the protection against harmful physical acts is greater than against offensive speech.[140] For example, while the ECtHR has concluded that all instances of sexual violence cause harm to the sexual autonomy of individuals, a distinction has been made between physical acts—such as rape—and, for example, the non-consensual recording of nudity. The former, involving physical harm, is included within the scope of the prohibition on torture, inhuman or

[135] *Perrin v the United Kingdom* App no 5446/03 (ECtHR, 18 October 2005). Both the UN Human Rights Committee and the ECtHR have, however, rejected the proposition that public morality *per se* justifies an intrusion on an individual's sexuality, as it must be balanced against the degree of intimacy to the person's identity. *See*, for example, *Toonen v Australia*, Communication No. 488/1992, UNHRC (31 March 1994) UN Doc CCPR/C/50/D/488/1992; *Dudgeon v the United Kingdom* (1981) 4 EHRR 149.

[136] *See*, for example, *Alekseyev v Russia* App nos 4916/07, 25,924/08 and 14,599/09 (ECtHR, 21 October 2010), paras. 85–86.

[137] Including incitement to racial discrimination (e.g. Art. 4 ICERD) and sexual harassment (Art. 40 Istanbul Convention).

[138] White (2006), p. 73.

[139] This is further explored in Sect. 3.3.4.

[140] *Söderman v Sweden* (2014) 58 EHRR 36.

degrading treatment and the latter within the right to privacy.[141] Although a gradation of harm is not *per se* inimical to human rights law, this hierarchy has gendered consequences. For example, a UN Secretary General report on violence against women notes that emotional and psychological aspects of abuse tend to be neglected.[142]

In terms of hierarchy, certain forms of speech are also considered more harmful than others in international human rights law, further discussed in relation to the freedom of expression. State *obligations* to combat hate speech or exclude it from protection are included in a range of treaties.[143] Furthermore, obligations to ensure liability for Internet media publishers has solely been affirmed in relation to hate speech and not, for example, defamation and harassment.[144] In contrast, states are *allowed* to restrict speech on the basis of public morals, such as pornographic material.

Similarly, the issue of causality in international human rights law impedes the recognition of certain forms of speech-based harm. Causation concerns the nexus between the harm experienced by the applicant and the alleged act or omission by the state.[145] While direct causality may not be required to the same extent and is not as conceptually developed in international human rights law as in domestic law, certain standards in relation to both individual and group-based harm may be evinced. For example, in certain individual cases of the ECtHR, the establishment of a nexus has been briefly touched upon.[146] The nexus depends on the right and the type of obligation. At times the ECtHR refers to 'objective scientific research'[147] or reports at the domestic or international level.[148] Causality is particularly difficult to verify in relation to omissions, that is, an aspect of positive obligations, and is thus often speculative.[149] In relation to this category of obligations, the ECtHR does not apply a "but for" test, that is, that but for the omission by the state the harm would not have ensued.[150] Rather, the Court has considered that '[a] failure to take reasonably available measures which could have had a real prospect of altering the outcome or mitigating the harm is sufficient to engage the responsibility of the State'.[151]

[141] ibid.

[142] UNGA, 'In-Depth Study on all Forms of Violence against Women: Report of the Secretary-General' (6 July 2006) UN Doc. A/61/122/Add.1, para. 117.

[143] Art. 4 ICERD (obligation to prohibit); Art. 17 of the ECHR (exclusion from protection).

[144] *Delfi v Estonia* (2014) 58 EHRR 29.

[145] Stoyanova (2018), p. 311.

[146] *L.C.B. v the United Kingdom* (1999) 27 EHRR 212.

[147] *Brincat and Others v Malta* App nos 60908/11, 62110/11, 62129/11, 62312/11 and 62338/11 (ECtHR, 24 July 2016), para 106.

[148] *Opuz v Turkey* (2010) 50 EHRR 28.

[149] Stoyanova (2018), p. 312.

[150] *E. and Others v the United Kingdom* (2003) 36 EHRR 31, para 99.

[151] *O'Keeffe v Ireland* (2015) 59 EHRR 15, para. 149; *Opuz v Turkey* (ECtHR), para. 136.

However, this is not applied consistently.[152] Since positive obligations require that states adopt certain measures, the non-existence of a law may be sufficient evidence of a violation, that is, causality may not be directly proven but is implicit. At the same time, positive obligations require the adoption of "effective" measures, entailing that a determination of cause and effect is relevant, in order to ensure that they fulfil their aim.[153]

In terms of causality to inequality, similar to linguistic philosophy, the theoretical distinction between speech and acts is not rigid in international human rights law. The content of "expression" protected by the freedom of speech has been determined on an *ad hoc* basis. Broadly, it covers all forms of communication where a message from one person is conveyed to another.[154] For example, according to the jurisprudence of the ECtHR, speech includes not only spoken or written words, but also paintings,[155] photographs,[156] symbols[157] and, in certain instances, clothes[158] and conduct[159] intended to convey an idea or information. Similarly, public nudity—arguably for the purpose of conveying an opinion on the inoffensive nature of the human body—has been considered a form of expression.[160] In contrast, speech has seldom been categorised as a form of act. Even in cases on hate speech and obscenity, regional human rights law courts and UN treaty bodies rarely evaluate such speech from the viewpoint of linguistic philosophy, although individual judges of the ECtHR have argued that hate speech constitutes acts.[161] The nature of speech does thus not remove its status as speech. In fact, deeply offensive, shocking or disturbing speech is also encompassed, although it may be restricted.[162]

Due to the structure of this legal regime, with its premier enforcement mechanism through individual complaints before international courts and treaty bodies, as well as a narrow approach to discrimination, the impact of speech on equality is rarely and

[152] *See* Stoyanova (2018) for a general overview. At times, the Court considers whether harm is 'caused', 'attributable' or 'imputable' to the state. *See*, for example, *Dodov v Bulgaria* App no 59548/00 (ECtHR, 17 January 2008) ('direct causal link'); *Fadeyeva v Russia* (2007) 45 EHRR 10, para. 92 ('sufficient nexus'); *E. and Others v the United Kingdom* (ECtHR), para. 100 ('significant influence on the course of events').

[153] For example, *M. C. v Bulgaria* (2005) 40 EHRR 20 para. 150; *Maslova and Nalbandov v Russia* (2009) 48 EHRR 37, para. 91.

[154] UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)' (2011) UN Doc. CCPR/C/GC/34, para. 11.

[155] *Müller and Others v Switzerland* (1988) 13 EHRR 212.

[156] *Von Hannover v Germany* (2005) 40 EHRR 1.

[157] *Donaldson v the United Kingdom* App no 56975/09 (ECtHR, 25 January 2011), para. 20.

[158] *Stevens v the United Kingdom* App no 11674/8 (ECtHR, 3 March 1896).

[159] Such as demonstrations. *See Steel and Others v the United Kingdom* (1998) 28 EHRR 603.

[160] *Gough v the United Kingdom* App no 49327/11 (ECtHR, 28 October 2014), para. 150.

[161] *Vejdeland and Others v Sweden* (2014) 58 EHRR 15, Concurring opinion of Judge Yudivska joined by Judge Villiger; Tulkens (2012), p. 281.

[162] UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression' (2011), para. 11; *Handyside v the United Kingdom* (1976) 1 EHRR 737, para. 49.

inconsistently addressed. This comports with the liberalist approach, which is more inclined to prohibit individual instances of speech, for example, incitement of imminent lawless action, as opposed to a communitarian viewpoint.[163] Whether equality may be harmed is also contested in international human rights law.[164] As equality is a norm rarely applied in the adjudication of individual cases, the standard of evaluation is unclear. However, international human rights law is becoming increasingly empirical, for instance, viewed in the statistical assessment of the compliance of states with treaty regulations.[165] Harm is in this sense not assessed in relation to individual offences and causation is to a degree reduced through this approach.[166] Furthermore, indirect discrimination may be proven through statistics on the prevalence of an offence against a particular group. This entails that although it is easier to prove harm against a particular group through its widespread occurrence, it is more difficult to affirm a correlation between certain types of material—such as pornography or hateful comments—and inequality. Thus, although substantive equality may partly be measured quantitatively, questions such as *how* discriminatory attitudes and customs may be eliminated benefit from a theoretical anchoring.[167]

In practice, it appears that UN treaty bodies and regional human rights law courts in many cases involving the lawfulness of restrictions on the freedom of expression presume harm without a clear indication of support, particularly in relation to social or group-based harm. That is, whether this presumption is empirical or theoretical is not specified. For example, the ECtHR prohibits hate speech in order to protect individuals from the 'harmful effects of such expression'[168] and 'their potential to lead to harmful consequences'.[169] In *Aksu v Turkey*, it held that '. . .any negative stereotyping of a group, when it reaches a certain level, is capable of impacting on the group's sense of identity and the feelings of self-worth and self-confidence of members of the group'.[170] In this regard, the Court emphasised diversity as an aspect of genuine democracy, obliging states to protect certain groups from harm.[171] In relation to the prohibition on the incitement to violence, often in the form of hate speech, an aspect of the legal assessment is the probability of violence occurring. For example, the ECtHR has considered the manner in which the

[163] Massaro (1991), p. 235.

[164] Brison (1998), fn. 37.

[165] Arguably, '[a]s elsewhere, a field which was once defined almost exclusively by either doctrinal argument or normative prescription has been transformed into one preoccupied with measurement.' *See* McGrogan (2016), pp. 620–621.

[166] McGrogan (2016), p. 623.

[167] ibid., p. 624.

[168] *Vejdeland and Others v Sweden* (ECtHR), para. 44. *See* also the concurring opinion of Judge Yudivska joined by Judge Villiger, noting that statistics indicated an elevated risk of hate crimes and that hate speech creates a climate of hatred.

[169] *Stomakhin v Russia* App no 52273/07 (ECtHR, 9 May 2018), para. 93.

[170] *Aksu v Turkey* App nos 4149/04 and 41029/04 (ECtHR, 15 March 2012), para. 58.

[171] ibid., para. 44.

statements were made, and their capacity to—directly or indirectly—lead to harmful consequences.[172] Meanwhile, the European Commission has in relation to sexist hate speech noted that this '. . .*can* escalate into hate crime offline'.[173]

Similarly, the United Nations Committee on the Elimination of Discrimination against Women (CEDAW Committee) in General Recommendation No. 19 and the Beijing Declaration affirm the correlation between pornography and gender-based violence.[174] The CEDAW Committee has noted that media and advertisements are increasingly pornographic, focused on idealised body shapes and portray women as sex objects, which '*may*. . .contribute to the increasing problem of eating disorders among young women and girls'.[175] Negative gender stereotypes presumably give rise to '. . .intimate partner violence, a youth culture that is increasingly marked by the objectification and sexualisation of girls, and girls presenting themselves in a highly sexual manner'.[176] Furthermore, the Committee has affirmed that the existence of gender stereotypes in the media 'has an impact on educational choices and the sharing of family and domestic responsibilities between women and men' and is a root cause of sexual violence against women.[177] The Committee has also accepted domestic studies on the causality between pornography and sexual abuse, calling for further research on the impact of pornography on gender-based violence.[178] Similarly, the prevalence of pornography and the 'sexualization of the public sphere' in the State party '*may* exacerbate sexual harassment and gender-based violence against women and girls'.[179] The UN Human Rights Committee (UN HRC) has also held that pornography 'is *likely* to promote these kinds of treatment of women and girls'.[180] It appears that the causality between speech and such social harms is thus to a degree speculative, that is, that the material *may* or is *likely* to cause harm, and appears to be theoretical rather than empirical.

[172] *Perinçek v Switzerland* (ECtHR), para. 206.

[173] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', para. 22. Emphasis added.

[174] CEDAW, 'General Recommendation No. 19 on Violence Against Women' (1992) UN Doc A/47/38, para. 12; Beijing Declaration and Platform for Action, Report of the Fourth World Conference on Women, Beijing, 4–15 September 1995, UN Doc A/CONF.177/20 and UN Doc A/CONF.177/20/Add.1, Platform for Action, para. 118.

[175] CEDAW, 'Concluding Observations on the Seventh Periodic Report of Finland' (10 March 2014) UN Doc. CEDAW/C/FIN/CO/7, para. 14. Emphasis added.

[176] CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017) UN Doc. CEDAW/C/NOR/CO/9, para. 22 (c).

[177] CEDAW, 'Concluding Observations on the Combined Seventh and Eighth Periodic Reports of Japan (10 March 2016) UN Doc. CEDAW/C/JPN/CO/7–8, paras. 20 (a) and (c).

[178] CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017), paras. 22 (d) and 23 (c).

[179] CEDAW, 'Concluding observations on the Combined Seventh and Eighth Periodic Reports of France' (25 July 2016) UN Doc. CEDAW/C/FRA/CO/7–8, para. 18 (d). Emphasis added.

[180] UN HRC, 'General Comment No. 28: Article 3 (The Equality of Rights Between Men and Women)' (20 March 2000) UN Doc. CCPR/C/21/Rev.1/Add.10, para 22. Emphasis added.

Meanwhile, the consideration of gender in the assessment of harm at an individual level has in certain instances been employed in international human rights law. Whereas particular regional courts are predominantly gender neutral in their approach, such as the ECtHR, other institutions have considered gender as a relevant factor. An example is the *Miguel Castro-Castro Prison* case of the Inter-American Court of Human Rights (IACtHR), concerning forced nudity of men and women.[181] Due to the higher risk for women of being subjected to sexual violence in general, forcing women in prison to remain nude was deemed to cause enhanced levels of psychological harm in comparison to men, as a result of the presumed fear of further sexual abuse. The individual harm was thus considered in relation to a heightened risk of harm at a social level. The Inter-American Commission on Human Rights (IACmHR) has also affirmed that women may suffer more severe consequences of sexual violence than men, for example, owing to social stigma, which engenders additional psychological suffering.[182] Although the harm may arise as a result of harmful stereotypes on, for instance, gender roles and sexuality, the view of the IACtHR is that since it causes actual harm, it must be remedied.[183] This is aligned with the approach of the United Nations Convention of the Elimination of All Forms of Discrimination (CEDAW) and the regional women's rights conventions, that women suffer exclusive or disproportionate harm in relation to certain offences.[184]

As for online speech, the heightened risk of harm on the Internet has been noted by UN treaty bodies, the CoE and the ECtHR. For example, in *Editorial Board of Pravoye Delo and Shtekel v Ukraine*, the ECtHR held that '[t]he risk of harm posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press'.[185] Particularly anonymity and the permanency of speech was noted as factors exacerbating the harm of speech.[186] The

[181] *Case of Miguel Castro-Castro Prison v Peru* (merits, reparations and costs) IACtHR Series C No. 160 (25 November 2006), para. 306.

[182] *Fernandez Ortega* et al. *v Mexico* (preliminary objections, merits, reparations, and costs) IACtHR Series C No 224 (30 August 2010), para. 124; *Rosendo Cantú* et al. *v Mexico* (Preliminary objections, merits, reparations, and costs) IACtHR Series C No. 216 (31 August 2010), para. 114.

[183] *Case of Artavia Murillo* et al. *v Costa Rica* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 257 (28 November 2012), para. 302.

[184] Inter-American Convention on the Prevention, Punishment and Eradication of Violence Against Women (Belém do Pará Convention) (1994), 33 i.l.m. 1534 (1994); Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (the Istanbul Convention) (2011), CETS No. 210, entered into force 1 August 2014; Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (Maputo Protocol) (2000), Adopted by the second Ordinary Session of the Assembly of the Union, CAB/LEG/66.6; entered into force 25 November 2005 and CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19′ (14 July 2017) UN Doc. CEDAW/C/GC/35.

[185] *Editorial Board of Pravoye Delo and Shtekel v Ukraine* App no 33014/05 (ECtHR, 5 May 2011), para. 63.

[186] *Delfi v Estonia* (ECtHR), para. 147.

UN Special Rapporteur on Violence against Women and GREVIO have also noted the aggravating effect of the Internet on gender-based violations and materials affirming gender stereotypes.[187] Nevertheless, the general standpoint appears to be that, in view of Internet architecture, online speech should not be more or less protected than in other fora, despite the fact that the harm of online speech is potentially greater.[188] When assessing harm, the ECtHR has in fact taken into consideration the frequently low-register style of language on the Internet, as a factor reducing the harmful impact of abusive speech.[189] The UN Special Rapporteur on the Freedom of Expression also notes that sanctions applied to offline defamation may be disproportionate online, given the ability of the individual concerned to exercise his/her right to instantly reply to restore harm.[190]

3.2.3.4 Conclusion

Gender-based violations on the Internet involve both individual offences and harm against women as a group. Individual instances of gendered offences on the Internet take the form of both speech and physical acts, including sexual violence, stalking, harassment and defamation, with detrimental physical, psychological and economic consequences. The effects are also group-based in relation to such individual offences, by disproportionately affecting women and thus contributing to gender inequality and the undermining of democracy. Meanwhile, group-based harm, negatively affecting the social status and security of women, includes sexist hate speech and pornography objectifying women. Collectively, both individual and group-based forms of harm generate debilitating gender stereotypes. However, it is clear that the concept of harm and its assessment at both the domestic and international level impede the recognition of gender-based online offences in several regards.

The harm principle employed primarily at the domestic level, but also reflected in international human rights law, requires the demonstration of a risk of harm to individuals in order for the state to prohibit conduct. This combines generalisations of harm and subjective experiences. The former can be evinced from empirical

[187] UNHRC, '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); GREVIO, 'General Recommendation No. 1 on the digital dimension of violence against women'.

[188] *M.L. and W.W. v Germany* App nos 60798/10 and 65599/10 (ECtHR, 28 June 2018), para. 113. *See* also Stone (2010), p. 175.

[189] *Payam Tamiz v the United Kingdom* App no 3877/14 (ECtHR, 19 September 2017), para. 81; *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* App no 22947/13 (ECtHR, 2 February 2016), para. 77.

[190] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27, para. 27.

studies but also theoretical frameworks, while subjective experiences concern the particular consequences in the individual case. Stemming from liberalist theories, the harm principle at the domestic level of many states primarily encompasses physical injury. In instances where psychological harm is recognised, it is generally considered less harmful. This is linked to a perceived lesser severity, the indeterminacy of objective hurt and issues of causality between speech and harm. Causality also presents a challenge when assessing group-based injuries and has thus mainly been advanced through theoretical discourses. This approach to harm thus inhibits the recognition of speech-based offences on the Internet where the injury is primarily psychological.

International human rights law does not explicitly employ the harm principle. Nevertheless, elements of the principle in legal philosophy are evident also in this body of law. There is also an interplay between domestic and international human rights law on theories of harm of particular offences, such as sexual harassment and pornography, where views on harm at the national level have influenced the definition and scope of rights.[191] From the viewpoint of international human rights law, formulating the harm of various forms of speech or acts is relevant in order to assess both *whether* they transgress provisions and *which* rights are violated. In turn, it may affect balancing exercises in instances of competing interests. Gradation of harm is embedded in the international human rights law regime, in the hierarchy of rights and the scope of state obligations, with speech-based and online offences generally not considered as severe as physical acts or conduct IRL. Although formulations of causality between harm and acts/omissions of states are tentatively made in individual cases, such are at best theoretical and at times speculative in relation to group-based injury, for example, concerning harmful pornography. International human rights law would as such benefit from a more pronounced theoretical perspective, from the viewpoint of substantive equality.

Meanwhile, feminist theories and linguistic philosophy provide arguments for theorising harm to gender equality. The predominant approach to harm has been criticised from a feminist perspective as being male-oriented. Gender-based injury *per se*, and in particular online harm, is in many instances not considered sufficiently grave to rise to the level of a violation and/or crime at the domestic level. The liberalist formulation of the harm principle identifies the injury independently from the context in which it occurs and the identity of the victim. In contrast, feminist scholars consider that context informs the degree and form of harm and that a failure to bear in mind social gender hierarchies leads to an eschewed understanding of harm. In this regard, the question of whether gender has an impact on harm is both empirical and theoretical. As noted, there are gendered differences not only in relation to the disproportionate number of female victims and the form of

[191] For example, the approach that the harm of sexual harassment involves gender inequality in the workplace has been transposed from domestic to international law. The delineation of the harm of pornography developed by radical feminists such as Catharine MacKinnon has been tentatively adopted by the CEDAW Committee. *See* Sects. 4.3.2 and 4.5.

harassment, but also in how acts are experienced. Meanwhile, linguistic philosophy categorises speech in relation to their effects, including discrimination, which also provides an argument for considering how speech may constitute or cause, for example, subordination.

The fact that speech and conduct transpire *online* also affects legal assessments of harm. The Internet is generally viewed as amplifying the harm of speech, for example, reputational harm. Simultaneously, the threshold for what is considered harmful has been raised by, for example, the ECtHR.[192] According to the Court, the common use of offensive language and Wild West norms in this medium reduces the impact of speech.[193] This not only undermines the online/offline coherence of rights but also highlights that a contextual approach must not solely consider the forum but also the gendered effects of rights interpretation.

3.3 The Scope of Rights Online and Offline: Harm, Values and Concepts

3.3.1 Introduction

Whereas the previous section approached harm from a general standpoint, the following part provides an overview of theoretical and legal frameworks *vis-à-vis* the content of the freedom of expression and the right to privacy and considers their application to the Internet. Given the lack of codification of gender-based cyber offences, primarily these provisions are relevant.[194] The focus lies on the values and scope of rights. The values and aims of protecting a particular right—developed in legal and philosophical theories—provide further insight into what constitutes a set-back of interests, that is, harm, in accordance with Feinberg's taxonomy and what may/should be protected and/or prohibited by the state. Meanwhile, this is concretised in international human rights law. The approach to harm and the values of rights in turn affect balancing exercises and proportionality assessments, which is addressed at a general level. For example, both rights theories and international human rights law affirm a hierarchy of harm in terms of speech/physical acts as well as different forms of speech. These aspects—the values and content of rights, including hierarchies of harm, and the outcome of balancing exercises—may be affected by the transposition of rights to the online sphere, thus impacting on the possibilities of regulating gender-based online offences. For instance, with the Internet as the context for speech and acts, the protection against gender-based harm may be overridden by other interests. Additionally, certain concepts and

[192] Rowbottom (2012), p. 376.

[193] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 77.

[194] Certain offences discussed in Chap. 4 also fall within the scope of the prohibition on torture, inhuman or degrading treatment.

legal assessments associated with the two rights may be affected by Internet architecture and cyber norms. This overview will serve as a background to the more specific analysis of gender-based violations in Chap. 4.

3.3.2 The Freedom of Expression

3.3.2.1 Rights Theories

As previously noted, many of the gender-based violations discussed in this book are conducted through speech. This includes harassment, the non-consensual dissemination of photographs or personal information, defamation and sexist hate speech. Pornography also constitutes a form of expression. The freedom of expression of the speaker is thus balanced against the protection of women against harmful speech. Simultaneously, the freedom of expression of women is affected by such offences, as harmful speech may silence women's voices and opinions in the virtual sphere. As already mentioned, although the freedom of expression and the right to privacy are of equal value, liberalism in many instances considers the freedom of expression to be of particular importance. There are multiple reasons for this, both explicit and implicit. Some are practical[195] and others ideological, evident through rights theories. The main arguments for protecting speech, in certain instances even when causing harm to others, can broadly be categorised as the discovery of truth (a marketplace of ideas), individual autonomy and democratic self-governance.[196] There is thus a distinction between the view of the freedom of expression as an intrinsic value to the individual, as opposed to one of instrumental social value, such as to further democracy and self-governance. International human rights law—and Internet architecture—mainly embody the latter two theories, which are thus discussed in the following.

The value of the freedom of expression to individual autonomy stems from the liberalist ideology. The protection of personal choice is key, with the expression of individual ideas considered outside the realm of state coercion or intrusion.[197] Accordingly, a person cannot develop morally or intellectually unless a person is

[195] Speech, as opposed to conduct, is more likely to be over-controlled by the state. *See* Schauer (1993), p. 639; Alexander and Horton (1983–1984), p. 1328.

[196] The theoretical concept of a "marketplace of ideas" as the foundation of the freedom of expression considers that the expression of opinions must be protected as a necessary or useful means of discovering the truth. The theory is primarily associated with ideas developed by John Stuart Mill, although introduced by Justice Oliver Wendell Holmes of the US Supreme Court. *See* Mill (1859), pp. 46–47; *Abrams v United States*, 250 U.S. 616 (1919) dissent by Justice Oliver Wendell Holmes Jr. *See* also discussion in Leiter (2010), p. 163; Levmore and Nussbaum (2010), p. 8.

[197] Chander (2010), p. 126.

allowed to express his/her views and debate them with others.[198] The free formulation of ideas and opinions leads to psychological benefits and to more informed choices and independent judgments.[199] The theory thus emphasises the intrinsic value of expressing oneself, leading to self-development and the enhancement of decision-making capabilities, regardless of the content of the ideas expressed.[200] Although the theory on the freedom of expression from the viewpoint of autonomy mainly focuses on the interests of the speaker, it may also extend to the audience. A protectionist approach by the state, limiting speech deemed harmful, is considered to disregard the autonomy also of members of the audience.[201] However, such a categorical approach to audience rights is considered impracticable.[202]

If speech is viewed as beneficial to the individual speaker *per se*, there is less room for state interference, even when causing harm. Nevertheless, the autonomy approach does not necessarily involve an unlimited right to free speech. It may allow for restrictions on speech that diminishes the autonomy of others, such as insulting or intimidating speech, for example, applicable to hate speech.[203] Accordingly, '[s]elf-expression should receive little protection if its sole purpose is to extinguish the self-expression of another'.[204] As such, two aspects of the protection of autonomy may conflict. Non-restraint may in fact have gendered consequences by increasing the prevalence of harmful speech particularly against women, evident online. Although the protection of individual autonomy is essential in ensuring the recognition of gender-based offences as violations of international human rights law, such as sexual autonomy, the application of the theory in relation to the freedom of expression has thus been criticised from a feminist perspective. Whereas the autonomy rationale relies on Kantian ideals, with the freedom of expression based on the presumption of individuals as free, rational and autonomous beings, it arguably discounts social realities and contexts, with women commonly constrained in their autonomy and in their access to public spheres of debate.[205]

Meanwhile, the theory on the instrumental value of speech in safeguarding democracy has primarily been developed by Alexander Meiklejohn.[206] It focuses on the benefits of the freedom of expression to society at large rather than its individual value. In order for people to be self-governing and able to perform

[198] Scanlon (1979).

[199] Levmore and Nussbaum (2010), p. 8.

[200] Moon (1985), p. 344.

[201] Scanlon (1979).

[202] For example, protection against false advertisements that individuals cannot verify is generally considered necessary for the protection of public health, thus understood as an overriding interest. *See* Barendt (2005), p. 406.

[203] Brown (2015), p. 60.

[204] Citron (2010), p. 46.

[205] MacKinnon (1993), p. 78.

[206] Meiklejohn (1948). Similarly, theories on deliberative democracy developed by John Rawls and Jürgen Habermas press the importance of free speech for citizen participation and thus for sound political decisions. *See* Rawls (1993); Habermas (1996).

essential tasks in the political process, they must be able to access information necessary for decision-making. It places positive obligations on states to provide citizens with tools to become better informed. What is essential is not that everyone speaks, but that everything worth saying is made public.[207] A free exchange of a wide diversity of opinions is necessary so that people receive information useful for meaningful participation in governance. In some respects, this overlaps with the theory on speech being essential to individual autonomy. In order for democracy to function, individuals must be informed and given the opportunity to develop decisions autonomously. As a consequence, they must be made aware of a variety of beliefs and have access to relevant information, which is in the interest of the speaker as well as the audience.[208] Nevertheless, although the aim is to ensure a plurality of opinions, a right to speak is not necessarily construed as essential to each individual.

Meiklejohn's theory is, for obvious reasons, focused on protecting political speech, in order to deter the potential abuse of power of public officials. However, the content of political speech and the understanding of democracy are not clearly delineated. Even though there are different approaches to the concept of democracy, in part based on the historical and cultural context, it encompasses two core factors: (1) popular sovereignty and (2) the right of each citizen to participate in the process by which society's decisions are made.[209] What is then speech conducive to democracy? This can be approached in a broad or narrow sense. For example, Meiklejohn argued that the development of the intellect, integrity and sensitivity of individuals aids democratic self-governance, encompassing a full range of communication from which the voter derives 'the capacity for sane and objective judgment', relevant for wider social and moral debates.[210] This includes educational, artistic and scientific expression,[211] thus bordering on autonomy rationales. Similarly, democratic culture may be considered an aspect of democracy, providing a broader view of valuable speech, including art, gossip and parody.[212] Accordingly, democracy must not be approached narrowly to solely concern debate on public issues or elections.[213] It encompasses also social movements, which may influence institutions, practices and customs.[214] Nevertheless, ambiguity remains as to what constitutes political speech and speech in the public interest in a democracy.

[207] Meiklejohn (1948), p. 25.
[208] Meiklejohn (1948), p. 25; White (2006), p. 61.
[209] Weinstein (2009), p. 25.
[210] Meiklejohn (1961), p. 256.
[211] ibid., p. 257.
[212] White (2006), p. 62.
[213] Balkin (2004), p. 2.
[214] ibid., p. 37.

3.3.2.2 International Human Rights Law

The protective interests and aims of the freedom of expression in international human rights law are generally not explicit, be it in treaties, case law or soft law. However, when noted, it appears that an amalgam of values inspires its protection. International human rights law courts and bodies often refer to both the intrinsic and instrumental values of speech. The duality of the protective interests as values to the speaker (self-fulfilment) and to the audience (democracy), is noticeable in that the freedom of expression engages both a right to express oneself and for individuals to access information. The freedom of expression thus broadly consists of two aspects: an individual dimension and a collective dimension.[215] Awareness of other people's opinions and information is considered as important as the right to impart your own.[216]

For example, the UN HRC in General Comment No. 34 provides that the values underlying the freedom of expression are multiple: it is considered a fundamental condition for the full development of the person,[217] essential in a democratic society[218] and necessary for the realisation of transparency and accountability, vital for the protection of human rights.[219] The ECtHR has similarly emphasised the value of the freedom of expression to individual autonomy. The Court in *Handyside v the United Kingdom* and *Lingens v Austria* affirmed that Article 10 'constitutes one of the essential foundations. . .for each individual's self-fulfilment',[220] and one of the basic conditions 'for the development of every man'. [221] The value of expression to autonomy also to an extent encompasses non-truthful statements, that is, the right is considered that of the speaker.[222] Nonetheless, the protection of autonomy appears as a more minor consideration than the value to democracy, with the ECtHR in practice mainly considering whether speech

[215] For example, the IACtHR has affirmed that: '[i]t requires, on the one hand, that no one be arbitrarily limited or impeded in expressing his own thoughts. In that sense, it is a right that belongs to each individual. Its second aspect, on the other hand, implies a collective right to receive any information whatsoever and to have access to the thoughts expressed by others'. *See Ivcher-Bronstein v Peru* (merits, reparations and costs) IACtHR Series C No. 74 (6 February 2001), para. 146.

[216] *Herrera-Ulloa v Costa Rica* (preliminary objections, merits, reparations and costs) IACtHR Series C No. 107 (2 July 2004), para. 110.

[217] UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression' (2011), para. 2.

[218] ibid., para. 2.

[219] ibid., para. 3.

[220] *Lingens v Austria* (1986) 8 EHRR 407, para. 41.

[221] *Handyside v the United Kingdom* (ECtHR), para. 49. *See* also *Mouvement Raëlien Suisse v Switzerland* App no 16354/06 (ECtHR, 13 January 2011), para. 49.

[222] *Salov v Ukraine* (2007) 45 EHRR 51, para. 113. At the same time, the Court has upheld prohibitions on Holocaust denial, on the basis of its demeaning character, which in part stems from its falseness. *See*, for example, *Pastörs v Germany* App no 55225/14 (ECtHR, 3 October 2019), para. 48.

contributes to the public interest or debate, that is, its social function.[223] In *Handyside v the United Kingdom*, in addition to noting the value to autonomy, the Court emphasised that the 'freedom of expression constitutes one of the essential foundations of [a democratic] society, one of the basic conditions for its progress . . .'.[224] Moreover, the IACtHR has affirmed its close relationship to democracy, holding that the '[f]reedom of expression is a cornerstone upon which the very existence of a democratic society rests.'[225]

In view of the aim of democracy, the IACmHR/IACtHR, the CERD and the UN Special Rapporteur on the Freedom of Expression have affirmed that the right to equality and the freedom of expression are complementary, especially in guaranteeing the right to equality of members of groups that historically have suffered from discrimination.[226] This freedom is thus considered essential for vulnerable groups to seek a balance in power, such as deconstructing stereotypes and offering alternative viewpoints.[227] The values of pluralism and diversity of information are reduced if certain groups are excluded from public debate. As disadvantaged groups have more limited institutional and private channels for expressing their views or to receive information, it similarly deprives society of their viewpoints. The effect is the silencing of certain groups, which become more vulnerable to marginalisation.[228]

[223] *See* discussion on the ECtHR and the freedom of expression as mainly democratic in Fenwick and Philipson (2003), p. 905.

[224] *Handyside v the United Kingdom* (ECtHR), para. 49.

[225] *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism (Arts. 13 and 29 American Convention on Human Rights),* Advisory Opinion OC-5/85, IACtHR Series A No. 5 (13 November 1985), para. 70.

[226] IACHR, 'Annual Report of the Inter-American Commission on Human Right: Report of the Special Rapporteur for Freedom of Expression', Chapter III (Inter-American Legal Framework of the Right to Freedom of Expression) (30 December 2009) OEA/Ser.L/V/II. Doc. 51., para. 9; *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism* (IACtHR), para. 50; CERD, 'General Recommendation No. 35: Combating Racist Hate Speech' (26 September 2013) UN Doc. CERD/C/GC/35, paras. 39–42; 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, para. 4.

[227] IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas', Annual Report of the Office of the Special Rapporteur for Freedom of Expression: Annual Report of the Inter-American Commission on Human Rights Vol. II (31 December 2015) OEA/Ser.L/V/II, para. 219.

[228] IACHR, 'Annual Report of the Inter-American Commission on Human Right: Report of the Special Rapporteur for Freedom of Expression', Chapter III (Inter-American Legal Framework of the Right to Freedom of Expression) (30 December 2009), para. 35.

A similar approach is espoused by the ACmHPR[229] and the UN HRC.[230] The UN Special Rapporteur on the Freedom of Expression has specifically noted that online gender-based violence hampers equal access to the freedom of expression and constitutes a form of gendered censorship.[231] Furthermore, the ECtHR has emphasised that democracy demands respect for diversity and, in relation to the freedom of expression, places restrictions on the abuse of dominant positions.[232] As such, positive obligations ensue in terms of protecting minorities and people with unpopular views, since they are more vulnerable to victimisation.[233] The protection of pluralism specifically in the media, aligned with a participatory model of democracy, has also been affirmed in case law of the ECtHR.[234]

The emphasis on the safeguarding of democracy affects several elements when assessing the scope of the right. Expressions engender various levels of protection, for instance, depending on the identity of the speaker and the content and aim of the speech, with certain forms of speech considered outside the scope of protection. For example, the ECtHR has constructed a hierarchy of protected speech, with differing levels of protection depending on the content and context. Speech beneficial to democracy is ranked higher than speech that offends morals. Speech enjoying special protection includes journalistic and political speech, that is, matters of public interest, with a wider margin of appreciation for states to restrict obscene or commercial speech.[235] In this regard, the importance of the right of the audience to receive information is emphasised.[236] The ECtHR applies a broad approach to what is considered to be in the public interest, affirming that it does not solely extend to issues of politics.[237] However, in *Hannover v Germany*, the ECtHR made it clear

[229] *Media Rights Agenda and Constitutional Rights Project v Nigeria*, ACmHPR, Communication Nos. 105/93, 128/94, 130/94 and 152/96 (31 October 1998), para. 54 (on media plurality in general); ACmHPR, 'Declaration of Principles on Freedom of Expression and Access to Information in Africa', Draft issued by the Special Rapporteur on Freedom of Expression and Access to Information in Africa, for consultation with States and other Stakeholders, pursuant to Resolution 350 (ACHPR/Res.350 (EXT.OS/XX) 2016) of the African Commission on Human and Peoples' Rights (30 April 2019), para. 8.

[230] *Adimayo M. Aduayom, Sofianou T. Diasso and Yawo S. Dobou v Togo*, Communications Nos 422/1990, 423/1990 and 424/1990, UNHRC, UN Doc. CCPR/C/51/D/422/1990, 423/1990 and 424/1990 (12 July 1996), para. 7.4.

[231] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021) UN Doc. A/76/258, para. 17.

[232] *Beizaras and Levickas v Lithuania* App no 41288/15 (ECtHR, 14 January 2020), paras. 106–107; *Aksu v Turkey* (ECtHR), para. 44.

[233] *Beizaras and Levickas v Lithuania* (ECtHR), para. 108.

[234] *Manole v Moldova* App no 13963/02 (ECtHR, 17 September 2009), para. 95. *See* discussion in O'Connell (2020), pp. 116–119.

[235] *Lingens v Austria* (ECtHR), para. 42; *Handyside v the United Kingdom* (ECtHR), para. 47; *Casado Coca v Spain* App no 15450/89 (ECtHR, 24 February 1994), para. 55.

[236] *Jersild v Denmark* (1995) 19 EHRR 1, para. 31.

[237] *Thorgeirson v Iceland* (1992) 14 EHRR 843 para. 64; *Tonsbergs Blad AS and Haukom v Norway* App no 510/04 (ECtHR, 1 March 2007), para. 87: 'Whether or not a publication concerns an issue

that, in relation to the media, speech in the public interest is to be distinguished from information that interests the public, such as gossip.[238]

Linked to this, those who are able to inform the public more directly, such as the press and politicians, receive broader protection.[239] That is, the identity of the speaker also generates a hierarchy of protection, connected to the democratic value of speech. It is often emphasised that the press holds a 'vital role' as a public watchdog.[240] Similarly, the UN HRC has underlined the special protection of the press, given its ability to influence public debate and public opinions on matters of legitimate public concern.[241] However, in certain cases, such as *Steel and Morris v the United Kingdom*, the ECtHR has rejected the approach that non-journalists should be treated differently than journalists, arguing that there exists a strong public interest in enabling '. . .individuals outside the mainstream to contribute to the public debate by disseminating information and ideas on matters of general public interest such as health and the environment'.[242] Furthermore, the concept of "journalist" has not been defined and it is unclear whether it extends only to authorised journalists. "Citizen journalists" who report on current events, armed conflicts and human rights abuse are widespread on the Internet.[243] In fact, both the UN HRC and the ECtHR have recognised that bloggers and popular users of social media may act as "public watchdogs" in relation to the freedom of expression.[244] The UN Special Rapporteur on the Freedom of Expression has also highlighted the importance of this new form of journalism for a richer diversity of views and opinions, as well as its critical role as a watchdog in countries where the freedom of expression is limited.[245] With broader protection come certain responsibilities in relation to the material they post, following basic ethical standards.[246] This is also the case for politicians, as their speech

---

of public concern should depend on a broader assessment of the subject matter and the context of the publication.'

[238] *Von Hannover v Germany* (ECtHR), para. 110.

[239] *Dammann v Switzerland* App no 77551/01 (ECtHR, 25 April 2006), para. 51 (the press); *Faruk Temel v Turkey* App no 16853/05 (ECtHR, 1 February 2011), para. 55.

[240] *Jersild v Denmark* (ECtHR), para. 35.

[241] *Toktakunov v Kyrgysztan*, Communication No. 1470/2006, UNHRC, UN Doc. CCPR/C/101/D/1470/2006 (21 April 2011), para. 6.3.

[242] *Steel and Morris v the United Kingdom* (2005) 41 EHRR 22, para. 89.

[243] UNHRC, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue' (4 June 2012) UN Doc. A/HRC/20/17, para. 61.

[244] *Magyar Helsinki Bizottság v Hungary* App no 18030/11 (ECtHR, 8 November 2016), para. 168; UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 44.

[245] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (11 August 2010) UN Doc. A/65/284, para. 62.

[246] In *Stoll v Switzerland* (2008) 47 EHRR 59, para. 103, the ECtHR held that they must 'provide reliable and precise information in accordance with the ethics of journalism'; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (11 August 2010), para. 94.

may reach a wide audience and is of particular importance.[247] At the same time as political speech enjoys special protection, it should be noted that criticism of public figures is more acceptable than of private individuals. For example, politicians must accept a greater degree of tolerance *vis-à-vis* criticism against them, as they knowingly lay themselves open to scrutiny.[248] In sum, theories on the purpose of the protection of the freedom of expression have an impact on the hierarchy of protected speech, which affects the scope for state restrictions.

As for regulation, the main position in international human rights law is viewpoint neutrality.[249] The free flow of information is considered the norm, with restrictions on speech the exception and counter speech preferred.[250] Content regulation may lead to a slippery slope of state interference and have a chilling effect on valuable speech. In criminalising offensive speech, states may overregulate or suppress also valuable speech. Simultaneously, it is apparent that society cannot function with an absolute freedom of expression and a chilling effect on certain speech is thus inevitable.

States may restrict most forms of speech, provided there is a legitimate aim.[251] As noted above, in certain instances, states are *obliged* to prohibit speech, such as hate speech.[252] Such speech is not considered protected by the freedom of expression.[253] The structure for assessing state interference is similar in major international human rights law treaties and applies equally to online speech.[254] States must provide a legitimate aim for restricting the right, indicate the necessity of the interference in a democratic society and use the least restrictive means, that is, proportionality.[255] The reasons for allowing state interference must be based on the enumerated aims. The aims are generic to the qualified rights and not reflective of a particular theoretical standpoint *vis-à-vis* the purpose and values in protecting the freedom of expression. Likewise, case law is frequently remiss of theoretical perspectives in this assessment.[256] Legitimate aims include state interests in preserving public morals and the

[247] *Féret v Belgium* App no 15615/07 (ECtHR, 16 July 2009), para. 63.

[248] *Lingens v Austria* (ECtHR), para. 42; UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 38.

[249] IACmHR, 'Hate Speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas', para. 15.

[250] Gagliardone et al. for UNESCO (2015), p. 5; IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas', para. 48.

[251] For example, Art. 10 of the ECHR; Art. 19 (3) of the ICCPR; Art. 13 (2) of the ACHR.

[252] For example, Art. 4 of the CERD.

[253] Such as in Art. 17 of the ECHR.

[254] For example, Art. 10 of the ECHR; Art. 19 (3) of the ICCPR; Art. 13 (2) of the ACHR.

[255] "Necessity" entails that a restriction must be more than merely useful, reasonable or desirable. *See Sunday Times v the United Kingdom* (1979) 2 EHRR 245, para. 59.

[256] For example, in *Vejdeland and Others v Sweden* of the ECtHR, several separate opinions were issued due to the fact that the reasoning for allowing restrictions on the freedom of expression was not developed theoretically by the majority in the judgment. In comparison, in the US, the marketplace of idea theory entails that e.g. child pornography is deemed to be of no value in the

protection of the rights of others, such as children, who are considered to be a vulnerable group, requiring special protection.[257] For example, illegal material is not *per se* excluded from the right to receive and impart information. The ECtHR has affirmed that the sharing of copyright-protected materials through torrent files is included in the provision. However, the state has a right to restrict the distribution for the aim of protecting others, that is, to protect copyright.[258] Certain aims, such as "morality", generate a wide margin of appreciation for states in determining the necessity of state interference and the type of measure employed.[259] Thus, construing pornography as a matter of morality does not provide as compelling grounds for requiring regulation as, for example, protection of gender equality. Nevertheless, it is not sufficient to base restrictions of speech on vague notions of offensiveness or immorality.

As noted above, certain forms of speech generate explicit obligations of prohibition, such as racist speech and incitement to violence in the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), the Protocol to the Convention on Cybercrime, Concerning the Criminalization of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (Additional Protocol to the Budapest Convention) and the International Covenant on Civil and Political Rights (ICCPR). [260] This also pertains to direct and public incitement to

marketplace, that is, the detriment outweighs the benefits. *See New York v Ferber*, 458 U.S. 747, 774 (1982).

[257] Art. 10 of the ECHR; Art. 19 (3) of the ICCPR; Art. 13 (2) of the ACHR.

[258] *Neij and Sunde Kolmisoppi v Sweden* App no 40397/12 (ECtHR, 19 February 2013).

[259] *Handyside v the United Kingdom* (ECtHR), para. 48; *Leo Hertzberg* et al. *v Finland*, Communication No. 61/1979, UNHRC, UN Doc. CCPR/C/OP/1 at 124 (2 April 1982), para. 10.3, albeit rejected in UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para. 36, as a principle.

[260] Art. 4 of International Convention on the Elimination of All Forms of Racial Discrimination (1965), 21 December 1965, G.A. res. 2106 (xx), Annex, 20 UN GAOR Supp. (No. 14) at 47, UN Doc. A/6014 (1966), 660 UNTS 195, entered into force 4 January 1969; Art. 3 of the Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems, (ETS No. 189) 28 January 2003; Art. 20 of the International Covenant on Civil and Political Rights (1966) G.A. Res. 2200A(xxi), UN Doc. A/6316, 999 UNTS 171, entered into force 23 March 1976. It should be noted that positive state obligations also extend beyond criminal or civil legislation to operational measures to ensure that persons are protected from acts impairing the enjoyment of the freedom of expression. For example, in *Özgür Gündem v Turkey* (2001) 31 EHRR 49, para. 46, the ECtHR held that if there is a known and credible threat against a journalist, the state is under a duty to take steps to protect him/her from physical attacks. Positive obligations were also affirmed in *Von Hannover v Germany* (ECtHR), para. 72. *See* also UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 23.

genocide,[261] child abuse images[262] and incitement to terrorism,[263] provided in various international treaties. More generally, the ECtHR has stated that democracy requires the sanctioning of certain types of speech. In *Gündüz v Turkey*, it held that '[t]olerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle, it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance'.[264] The list of expressions engendering positive obligations of criminalisation is thus limited. It does not explicitly extend to gendered speech-based offences. However, as noted, gender-based violence is a broad concept and includes both physical and psychological forms of violence, such as sexual harassment, which may include online speech. The elimination of gender stereotypes is also required by the CEDAW.[265] Likewise, obligations to prohibit "sexist hate speech" and obscene pornography is increasingly advanced in international human rights law, although primarily in soft law documents.[266] Thus, the list of speech engendering restrictions may certainly expand beyond explicit treaty provisions, to include various forms of gender-based offences.

In the main, the ECtHR does not approach harmful speech from the standpoint of obligations to prohibit, but in terms of exclusion from protection. Accordingly, expressions that conflict with the values proclaimed and guaranteed by the European Convention on the Protection of Human Rights and Fundamental Freedoms (ECHR), including democratic principles—that is, which attack or seek to

[261] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (10 August 2011) UN Doc. A/HRC/66/290, para. 23, relying on the statutes of the ICTY/ICTR and the Rome Statute. The Special Rapporteur for Freedom of Expression of the IACmHR has likewise affirmed that 'clearly illegal content or speech that is not covered by the right to freedom of expression (such as war propaganda and hate speech inciting violence, direct and public incitement to genocide, and child pornography)' engages state obligations. *See* IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013) OEA/Ser.L./V/II. CIDH/RELE/INF.11/13, para. 85.

[262] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (10 August 2011), paras. 20–22, relying on the Optional Protocol to CRC; Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (2011) OJ L335/1; The Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (Lanzarote Convention) (ETS No. 201) 25 October 2007.

[263] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (10 August 2011), para. 32, relying on Security Council Resolution 1624 (2005).

[264] *Gündüz v Turkey* (2005) 41 EHRR 5, para. 40.

[265] Art. 5 of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) (1979), GA res. 34/180, 18 December 1979, 1249 UNTS 13, entered into force 3 September 1981.

[266] *See* Sects. 4.4 and 4.5.

undermine democracy—are not protected under the ECHR in accordance with Article 17, even when speech expresses political ideas.[267] This again correlates the protection of speech with its contribution to democracy. What this entails is decided on a case-by-case basis. It depends on how the concept of democracy is interpreted and could, for example, include speech impeding equal participation in debate.[268] Pluralism and a diversity of viewpoints are necessarily components of the right, which supports arguments for ensuring equality both in terms of access to the Internet as well as content. This has, for example, been affirmed by the CoE.[269] In practice, however, the provision has been limited to excluding hate speech and incitement to violence from protection.[270]

3.3.2.3 The Freedom of Expression and the Internet

3.3.2.3.1 General Applicability

The freedom of expression is neutral *vis-à-vis* the medium. This is evident, for instance, in Article 19 of the Universal Declaration of Human Rights (UDHR), which holds that the freedom of expression protects information and ideas '. . .through any media and regardless of frontiers' and Article 19 (2) of the ICCPR, in which protection extends to ideas expressed orally, in writing, in print or 'any other media of his choice'.[271] The UN HRC has affirmed that the right comprises all forms of audio-visual, electronic and Internet-based modes of

[267] *Refah Partisi (the Welfare Party) and Others v Turkey* (2003) 37 EHRR 1, para. 98. *See* also Art. 5 of the ICCPR.

[268] O'Connell (2020), p. 94.

[269] CoE, 'Recommendation CM/Rec(2013)1 of the Committee of Ministers to Member States on Gender Equality and Media (Adopted by the Committee of Ministers on 10 July 2013 at the 1176th meeting of the Ministers' Deputies)'. The CoE affirms the necessity of pluralism and diversity to fulfil democratic ideals. Accordingly, '[t]he media are centrally placed to shape society's perceptions, ideas, attitudes and behaviour' and should thus reflect the lives of men and women in all their diversity (preamble).

[270] For example, Art. 20 of the ICCPR and Art. 17 of the ECHR. *See* also *Gündüz v Turkey* (ECtHR), para. 41. The Court has also considered revisionism of the Holocaust as incompatible with democracy and human rights and a threat to public order. *See Garaudy v France* App no 65831/01 (ECtHR, 24 June 2003). This can be contrasted with the approach of the UN Human Rights Committee, which considers laws prohibiting Holocaust denial incompatible with the freedom of expression. *See* UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 49.

[271] Similarly, protection of the freedom of expression in media 'of all kinds' is mentioned in Art. 13 (1) of the American Convention on Human Rights (1969) O.A.S.T.S. 36, 1144 UNTS 123, entered into force 18 July 1978.

expression[272] and the ECtHR has reviewed a range of cases involving speech on news websites, social media and other online platforms.[273]

The freedom of expression not only involves the content of the information communicated online, but also the *means* through which it is disseminated, that is, the right to express an opinion in a forum of choice, that is, an access-oriented approach. For example, the IACtHR has held that the first dimension of the freedom of expression 'is not exhausted in the theoretical recognition of the right to speak or write, but also includes, inseparably, the right to use any appropriate method to disseminate ideas and allow them to reach the greatest number of persons'.[274] In this sense, the expression and distribution of ideas and information are indivisible, so that a restriction on the outlets for dissemination is a limitation on the freedom of expression.[275] Accordingly, the ECtHR in *Yildirim v Turkey* held that the creation and sharing of websites in a group run by Google Sites constituted a means of exercising the freedom of expression.[276] Video-hosting websites and mobile apps have also been affirmed as protected venues of expressing oneself.[277] As noted previously, states incur positive obligations to ensure accessibility to information and communication technologies (ICTs), particularly for vulnerable groups, that is, to 'create an enabling environment for all individuals to exercise their right to freedom of opinion and expression'.[278] Accordingly, the blocking and filtering of Internet websites has been criticised by the UN Special Rapporteur on the Freedom of Expression, regional human rights law bodies and various UN committees for impeding the freedom of expression.[279] Restricting access to the Internet *per se*, as an exceptional platform for transmitting and receiving ideas, or particular websites, thus limits the freedom of expression. Since cyber technologies offer an 'unprecedented capacity' for states and intermediaries to interfere with and control content

[272] UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 12.

[273] *See,* for example, *Delfi v Estonia* (ECtHR) and *Payam Tamiz v the United Kingdom* (ECtHR).

[274] *Ivcher-Bronstein v Peru* (IACtHR), para. 147

[275] *Herrera-Ulloa v Costa Rica* (IACtHR), para. 109.

[276] *Ahmet Yildirim v Turkey* App no. 3111/10 (ECtHR, 18 December 2012).

[277] *Cengiz and Others v Turkey* App no 48226/10 and 14027/11 (ECtHR, 1 December 2015) (video-service); *Magyar Kétfarkú Kutya Párt (MKKP) v Hungary* App no 201/17 (ECtHR, 20 January 2020) (mobile app).

[278] Including equal access, functional equivalence, accessibility, affordability and design for all. *See* UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (10 August 2011), para. 80.

[279] *Ahmet Yildirim v Turkey* App no. 3111/10 (ECtHR, 18 December 2012), UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), paras. 31, 75; IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 58; UN HRC Resolution 32/13 on the promotion, protection and enjoyment of human rights (1 July 2016).

online, much focus has been placed on such negative obligations in cases involving the Internet.[280]

Meanwhile, it is unlikely that the freedom of expression in the context of a right to access the Internet includes a right to express oneself on a particular website, as websites are mainly governed by private corporations and obligations may infringe on their property rights. As argued by the ECtHR, in a private sphere, positive obligations for states only arise where the bar on access to property effectively prevents the exercise of the freedom or destroys the essence of the right.[281] Similarly, the UN HRC has noted that individuals do not have access to all mediums of communication through the freedom of expression.[282] However, if the social networking market were to be monopolised by a few companies, a theoretical argument could be made that a right for the person to express him−/herself in a particular forum exists, if abiding by the terms of service of the company.[283]

At a general level, principles and obligations associated with the freedom of expression are the same online/offline. Accordingly, whereas the Internet is heralded as a particularly important forum to such values as democracy, it does not entail that the freedom of expression is more generous in this medium. Similar content-based restrictions on speech and state obligations are transposed to the Internet, that is, what is prohibited offline is also prohibited online. Although to a limited degree, positive obligations to protect individuals have, for example, been affirmed in relation to certain speech-based violations online, including domestic violence,[284] sexual violence[285] and hate speech.[286] In order to maintain the same values of rights, Internet architecture may thus not be used as an argument to lower individual protection against harmful speech. Nevertheless, in view of Internet design and cyber norms, several aspects of the freedom of expression may need to be revisited and, perhaps, revised.

The Internet affects the transposition of basic concepts integral to the freedom of expression, such as who is an author, journalist, publisher and victim. The Internet provides users with the ability to bypass the traditional gatekeepers of information and communication, such as publishers or broadcasters. Current gatekeepers are Internet Service Providers (ISPs), hosts and search engines. New technologies and social media have also created a shift in the role of individuals, from being merely

[280] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015) UN Doc. A/HRC/29/32, para. 1.

[281] An analogy can be drawn to the ECtHR case of *Appleby v the United Kingdom* (2003) 37 EHRR 38, involving protesters who were turned away from collecting signatures in a shopping centre and the question of whether there was an obligation for private property owners to physically allow people to express their opinions on their premises. In the case, there were alternative means of exercising the right, thus not engaging positive obligations for the state.

[282] *Leo Hertzberg* et al. *v Finland* (UN HRC), para. 10.2.

[283] Benedek and Kettemann (2013), p. 106.

[284] *Buturuga v Romania* App no 56867/15 (ECtHR, 11 February 2020).

[285] *K.U. v Finland* (2009) 48 EHRR 52.

[286] *Beizaras and Levickas v Lithuania* (ECtHR).

consumers of media to producers, creators and curators of information.[287] A recipient of information may simultaneously become a content provider through posting or re-posting comments. The swift changes between "publishing mode" to "private communication mode"—traditionally governed by different legal regimes—is one of the main challenges of Internet regulation.[288] Domestic and international courts are thus placed in the position of evaluating the need for symmetry between the rules of printed media and the Internet, such as transposing liability for publishers of newspapers to operators of websites and blogs. For example, traditional approaches to defamation hold the publisher liable. In cyberspace, this translates into the website host being liable for the content of comments of which it may have no knowledge. This has occasioned the ECtHR to determine the scope of secondary liability for Internet intermediaries and news websites, as well as the functions of hyperlinking and "liking" content, discussed further in relation to specific offences.[289]

Furthermore, the Internet has an impact on assessments of the intent and effect of speech, which is relevant in the categorisation of speech as specific offences, for example, whether speech constitutes harassment, defamation or incitement to violence, and the applicability of particular human rights law provisions. The difficulties associated with conducting linguistic analyses of online speech has been noted by the ECtHR, affecting the liability imposed on online media publishers and intermediaries.[290] With the absence of contextual indicators common in face-to-face communication, there is a higher risk of intent being misinterpreted. Online communication is mainly written and thus lacks such social cues as intonation and facial expressions that may indicate the intent of speech, for example, sarcasm. Character limitations on certain social media platforms also restrict the possibility of signaling intent.[291] In fact, the lack of tonal cues has led to the design of emoticons, emojis, gifs and abbreviations to convey the intention of the communication or to describe the content.[292] Such icons may be used to threaten and harass individuals.[293] These features thus provide a contextual nuance to the interpretation of the purpose and effect of speech and have been taken into account in defamation cases at the domestic level.[294]

However, symbols are also subject to misinterpretation, as they are inherently ambiguous and there may be no standard international meanings.[295] Studies indicate

[287] Gagliardone et al. for UNESCO (2015), p. 47.

[288] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, 'Illegal and Harmful Content on the Internet' (1996) 1996COM(96) 487, p. 8.

[289] *See* Sects. 3.4.3.2.3 and 4.3.4.

[290] *Delfi v Estonia* (ECtHR), para. 156.

[291] Lidsky (2018), p. 1910.

[292] ibid., p. 1907.

[293] Geneus (2018), p. 438.

[294] *Ghanam v John Does*, 303 Mich. App. 522, 545, 845 N.W.2d at 145.

[295] Lidsky (2018), p. 1885; Herring and Dainas (2020), p. 1.

that, for instance, emoji use and interpretation differ depending on the platform, culture and the relationship between the sender and the recipient.[296] Gender and age also affect the use of emojis and emoticons—more frequent among young females—and the purpose and tone of the particular emoji.[297] Meanwhile, online hate groups increasingly use steganography, symbolic markers or benign terms as code words for racial slurs in their online communication and the meaning of such icons must thus be understood in their context.[298] Additionally, online speech is frequently multimodal, for example, combining text, audio, video and graphics, which collectively convey meaning and its parts cannot be viewed in isolation. This adds complexity to legal assessments at the domestic and international levels, as well as monitoring by Internet intermediaries and automatic detection systems, which are generally text-based.[299] According to the UN Special Rapporteur on the Freedom of Expression, the failure of automated content moderation to assess cultural nuance may place women at particular risk, in view of transgressions of social and cultural norms depending on local contexts.[300]

A contextual approach must thus not only consider such online features but, arguably, also the jargon or discourse on a particular website or in subgroups on platforms, in view of the demographics of the audience.[301] Offensive speech in jest is more common on certain platforms, acceptable to particular audiences. In assessments of hate speech, defamation and the non-consensual disclosure of private information, the size of the audience is similarly an aspect that is taken into account when considering harmful effects. With a global audience, the potential harm of online publications is exacerbated. In certain instances, the risk of information "going viral" has been observed, for example, by the ECtHR.[302] As noted in Sect. 3.2 on the harm of online speech, the Internet *per se* has been viewed as both aggravating and reducing harm.[303] Nevertheless, mainly the webpage in question and the number of monthly visitors have been considered in practice, that is, the particular pocket of the Internet.[304]

The Internet also affects the level of control of speech and the audience. Speech is frequently altered, recontextualised, and appropriated as it travels from the initial

[296] Barbieri et al. (2016); Tigwell and Flatla (2016), p. 859; Herring and Dainas (2020), p. 2.

[297] Herring and Dainas (2020), p. 1.

[298] Siegel (2020), p. 57.

[299] Ullmann and Tomalin (2020), p. 78.

[300] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 81.

[301] Lidsky (2018), p. 1909.

[302] *Beizaras and Levickas v Lithuania* (ECtHR), para. 127.

[303] *Payam Tamiz v the United Kingdom* (ECtHR), para. 81*; Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 77.

[304] *Savva Terentyev v Russia* App no 10692/09 (ECtHR, 28 August 2018), para. 81; *Pihl v Sweden* App no. 74742/14 (ECtHR, 7 February 2017), para. 31.

speaker to the intended audience.[305] The ease with which information can be copied entails that the context of the original speaker may change. With a more limited ability to manage the audience, individuals have less control over the effects of online speech.

Given that contextual evaluations of the intent and effect of speech are more difficult online, it may require a turn to more objective elements of certain offences. This includes hate speech, which currently requires a consideration of the audience and the cultural context. Nevertheless, in relation to, for example, harassment and threats, linked to individual subjective experiences, such evaluations must necessarily remain, which requires a consideration of how the medium affects intent and effect.

3.3.2.3.2 Freedom of Speech Values

Whether and how the Internet undermines or safeguards the values of rights is also a factor in considering online/offline coherence. It affects the legitimacy of state constraints of speech, the view on liability (for the state, intermediaries and individuals), the content of rights (e.g. ensuring gender equality) and the importance placed on the freedom of expression on the Internet in the balance against other interests.

Internet architecture embodies and accentuates the two prominent values of the freedom of expression, to varying degrees. As noted above, from the viewpoint of cyberfeminism, the values embedded in technological design are a result of rational choice and not solely scientific inevitability. The ideology reflected in Internet architecture is principally libertarianism and its emphasis on individual autonomy. Individual users are provided a prominent role through its end-to-end layout, and are able to affect its design and content, as well as transmit or block information.[306] User anonymity in turn entails that social and legal restraints are limited, allowing individuals greater possibilities to participate in both harmless and harmful behaviour. In practice, this means that the freedom of expression is generally broad on the Internet. However, although several features of Internet architecture embody the value of individual autonomy in practice, international human rights law does not adhere to an unrestrained autonomy approach. Ensuring an online/offline coherence thus entails that the application of this right on the Internet cannot be more generous in terms of allowing speech-based offences, and effective means of preventing or responding to such speech must be developed in cooperation with intermediaries.

The value of the freedom of expression as a vehicle for democracy is also integral to the Internet. This forum promotes democratic ideals through its technical features, such as global accessibility to communicate and retrieve information, limited censorship by authorities and social decontextualisation, that is, anonymity obscuring social status cues which may otherwise limit participation. Minorities and

[305] Sellars (2016), p. 28.

[306] Karanasiou (2016).

marginalised individuals may, in theory, participate on an equal basis with others, with a focus on the content of communications rather than the identity of the speaker.[307] The lack of constraints associated with traditional media, such as the preference of expressions of the wealthy and the mainstreaming of mass media, provide individuals with possibilities to directly influence social and political events.[308] The Internet is particularly important from a speaker's perspective, since individuals are monitored in many other fora, be it in the press or broadcasting media.[309] The potential for an unprecedented sphere of democracy has thus been ascribed to the Internet by both cyberlibertarians and cyberfeminists. As mentioned, the significant value of the Internet may have an impact on the cost-benefit analysis in the balancing of rights. For example, the value to the public good may be perceived as greater than, at least, minor forms of individual harm, an approach implied by regional human rights law courts and commissions.[310] The democratic utility of the Internet is in such instances emphasised.

However, the uncritical acceptance of the democratic function of the Internet must be counterbalanced. Technological restraints limiting its democratic effect should also be recognised, such as filtering systems, the regulation of content by intermediaries and the customisation of information viewed by the user.[311] Although democracy generally allows for the dissemination of falsehoods, the widespread acceptance of misrepresentations online, for example, through cascade effects and group polarisation, may negatively affect the democratic process.[312] As such, it can be argued that the Internet does not guarantee autonomous agents of free choice but is characterised by a structure of interdependence similar to IRL.

Additionally, a theory which places an emphasis on ensuring democracy must consider the right to *equal* participation in the public sphere. That is, not only a consideration of the end result but also the process.[313] Although the democracy theory does not require that a state ensures that all individuals participate, the exclusion of certain social groups in public debate *per se* reduces the potential for a diverse political discourse. Public debate cannot be viewed as a process separate from socio-economic realities that constrain the ability of individuals to participate.

---

[307] Herring (1996), p. 477.

[308] Volokh (1995).

[309] Barendt (2016), p. 130.

[310] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 53; *Payam Tamiz v the United Kingdom* (ECtHR), para. 90.

[311] *See* Waldman (2013), pp. 348–349.

[312] Nussbaum (2010), p. 85. Research in psychology demonstrates that the Internet affects our brains in a manner that is not conducive to the development of arguments and the search for truth. Internet users tend to rely on conventional ideas and solutions rather than challenging such with independent viewpoints. *See* Carr (2010). However, it could be argued that certain features of Internet architecture balance these negative aspects, for example: (1) the sheer number of falsehoods may lead people to distrust information and become more critical of arguments and; (2) there is a greater possibility to instantly correct information perceived as false. *See* Nussbaum (2010), p. 104.

[313] Weinstein (2009), p. 27.

Bullying, harassment and hate speech deter valuable speech and prevent vulnerable people from participating in society as equals.[314] Meanwhile, the disproportionate level of harassment of women online limits their ability to conduct themselves as agents in the public domain. The male social dominance, also in the public sphere of the Internet, in effect silences female voices on the Internet.[315] It could thus be argued that an unregulated cyberspace is inimical to the ideals of a liberal democracy, as it may lead to discrimination and harm to minorities.[316]

In order to ensure the democratic value of equal participation, certain forms of speech thus warrant regulation.[317] In other words, it involves silencing certain forms of speech in order to facilitate speech by others. Empirical studies in fact indicate that regulation of harassment online does not have a chilling effect on the freedom of expression but overall encourages broader participation.[318] Similarly, it must be acknowledged that the proclaimed neutrality with which the state approaches speech on the Internet is in fact not neutral, but '. . .a value judgment that words [in this context] cannot cause harm or that the harm they do cause is not a social ill for the state or society to solve'.[319] As noted previously, by not restricting certain speech against vulnerable groups, the costs are borne by these groups in particular and may generate viewpoint discrimination as it undermines their ability to express their ideas. As an analogy, the case law of regional human rights law courts affirms that negligence (e.g. to investigate private acts of violence) may constitute indirect discrimination, for example, with regard to sexual or domestic violence, given the unequal effects on a particular group.[320] This creates an impetus to regulate harmful speech, in order to *increase* the democratic potential of the Internet.

3.3.2.4 Conclusion

In considering theoretical perspectives on the aims of protecting the freedom of expression, it is clear that the approach to restricting harmful speech varies. The aim of protecting speech to ensure individual autonomy views the moral and intellectual development of the person as integral, thus less involved with the content of speech. Meanwhile, the theory on democracy, connecting the freedom of expression to the value of democratic participation for the purpose of self-governance, is mainly concerned with the social benefits of speech. Different values are accordingly

[314] Chaffin (2008), p. 805; Levmore and Nussbaum (2010), p. 9; Citron (2010), p. 49; Murray (2016), p. 117.

[315] Megarry (2014), p. 53.

[316] Netanel (2000), p. 397.

[317] Brown (2015), p. 195.

[318] Citron and Penney (2019), p. 2329.

[319] Waldman (2013), p. 385.

[320] *Opuz v Turkey* (ECtHR); *González* et al. *(Cotton Field) v Mexico* (preliminary objections, merits, reparations and costs) (IACtHR), para. 164.

attached to certain categories of speech, based on their importance in relation to the aims, implying a hierarchy of protection.[321] This in turn has an impact on the assessment of the degree of harm required to restrict speech, as well as the scope and limitations considered appropriate. As discussed above, '[o]nly the incredible view that all expression, regardless of its subject or character, has value could sustain the idea that there is a significant clash. . .' between interests.[322]

Even though these theories allow for restrictions of speech, those rooted in individual autonomy—attaching an intrinsic value to expression—tend to provide the strongest protection of speech. If autonomy-based, the freedom of expression is considered a right mainly of the speaker, which entails a broader right to express offensive opinions.[323] From this perspective, solely speech which undermines the autonomy of another individual allows for restrictions, whereas broader possibilities exist according to other theories. If the freedom of expression relies mainly on the democratic ideal and is thus of instrumental value, political speech is considered of chief importance, thus allowing for a broader margin of appreciation for states to regulate other forms of speech. This includes harassment, which impedes equal participation in democratic debate. Nevertheless, difficulty arises in determining the value of particular forms of speech and the requisite level of harm for restricting speech. That is, what forms of speech vanquish individual autonomy? And which types of expression impede the democratic discourse? For example, it may be difficult to distinguish hate speech from political ideas. Similarly, depending on the argument, pornography projects political statements on sex/gender roles or is a vehicle for discrimination.[324]

Meanwhile, the Internet is, in many regards, beneficial to the values of the freedom of expression, such as enhancing democratic participation, access to information and by providing a platform for the self-expression of individuals. In fact, several of the core values of the freedom of expression are embedded in the architecture of the Internet. This pertains in particular to individual autonomy, with its end-to-end design, limited state regulation and broad possibilities for individuals to control its layout and content. The same technological features also embody the value of democracy, in the sense of democratic culture. However, these characteristics may also be inimical to the same ideals, with user anonymity and limited state interference allowing speech that vanquishes the autonomy of individuals to flourish. Furthermore, Internet architecture tends to affirm pre-existing ideas

[321] For example, in the US, the hierarchy may be simplified as: (1) political speech; (2) commercial speech; (3) obscenity and incitement to imminent lawless action. It is similar to the ECtHR, with a difference being the approach to hate speech (unprotected by the ECtHR). *See*, for example, an overview in Stone et al. (2020).

[322] Leiter (2010), p. 172.

[323] Accordingly; '. . .instrumentally based rights tend to be more fragile than morally based ones; they are vulnerable to being overridden . . . if the utilitarian calculus suggests that society would be better off without them. In contrast, rights that are justified in terms of moral rights of individuals as well as benefiting society as a whole tend to be sturdier. . .'. *See* Weinstein (1999), p. 14.

[324] *See* Sect. 4.5.

and acceptance of falsehoods, discourages in-depth reasoning, generates widespread low value speech, is plagued by a digital divide in terms of access and participation, and its content is increasingly mediated by intermediaries. This warrants a less idealistic approach to the Internet, and the recognition that the values of the freedom of expression must be viewed in light of this context, as its features affect the possibilities of ensuring the aims of this freedom. This may affect the evaluation of an online/offline coherence in relation to the freedom of expression.

Whereas several regional human rights law courts and UN treaty bodies have noted the value of the freedom of expression to individual autonomy, it appears that the main rationale in international human rights law is its contribution to democracy. The beneficial effects of the Internet to democratic participation are often emphasised by regional and international organisations and have served to limit the scope of state intervention. Given this connection between the Internet and the core value of the freedom of expression, it is clear that *a priori* importance is attached to this medium. Additionally, Internet architecture affects the scope of obligations in international human rights law, the evaluation of harm and the applicability of concepts and norms, that is, how rights are applied online. This has mainly engendered a more generous approach to the freedom of expression in this forum.

As stated above, feminist approaches to legal, political and linguistic theories on speech have mainly been reactive, that is, involved criticism of the main theories underlying the freedom of expression. It is clear that a variety of strategies have or may be employed from a feminist perspective to argue in favour of a regulation of gendered speech-based harm. An integral aspect is that rights and freedoms—such as the freedom of expression—must be interpreted in a manner that ensures substantive equality rather than upholding the power of the dominant group. Feminist critique includes the argument that certain material, such as hard-core pornography, does not constitute speech. It is rather considered a sexual aid for the purpose of sexual stimulation, without communicative intent.[325] Alternatively, certain speech may be categorised as harmful speech acts, including harassment, sexist hate speech and pornography, which silence and subordinate women or, alternatively, engender harmful effects, such as physical or mental harm. In relation to rights theories, it is argued that harmful gender-based speech does not advance the aims of the freedom of expression, be it individual autonomy (speech for the purpose of vanquishing the autonomy of another individual is not protected) or democracy (it restrains diversity in the political discourse). Online harassment limits the participation of women in democratic debate and undermines female autonomy. These forms of speech must thus be considered low value speech, if considered speech at all.

Several international bodies, including the United Nations Committee on the Elimination of Racial Discrimination (CERD), the IACmHR and the ECtHR have also emphasised the connection between the freedom of expression and substantive equality. By noting the broader instrumental values of the freedom of expression, it allows room for a gender-inclusive interpretation, accentuating the need for diversity

[325] White (2006), p. 62.

in participation and content, through restrictions on speech silencing sections of society. With the overall aim of substantive equality, speech in this sphere would not be treated in a more beneficial manner. As seen above, the concrete content of state obligations *vis-à-vis* the freedom of expression remains the same on the Internet, including non-interference in terms of monitoring and obligations to prohibit certain forms of speech. Naturally, care must be taken not to excessively restrain the freedom of expression, given the heightened risk of illicit censorship, which is also detrimental from a gender perspective. For example, repressive states may claim that content involving women's sexuality and reproduction is harmful and contrary to public morals.[326] However, where the application of rights require adaptation to the context, it is important that a functional equivalence is maintained, not lowering standards of protection, for example, in relation to gender-based harm.

At a more concrete level, the Internet also affects the interpretation of concepts—such as who is an author, journalist or publisher—and the ability of states or intermediaries to conduct linguistic assessments of intent, effect and the harm of speech. Such issues arise regardless of the form of speech but has particular consequences in relation to speech-based offences sensitive to nuance or context. This affects state obligations, including the regulation of intermediary liability and means of moderating content.

3.3.3 The Right to Privacy

3.3.3.1 Rights Theories and International Human Rights Law

In contrast to most international human rights, the right to privacy was not broadly included in domestic laws prior to international codification.[327] Its roots can be traced to Roman law, protecting individuals primarily against physical injury, but also certain forms of psychological harm.[328] However, it is often considered that the concept of privacy was first introduced in an academic article written in 1890 by Louis Brandeis and Samuel Warren, defining the right to privacy as a right to be let alone, that is, seclusion.[329] Given its *ad hoc* based regulation at the domestic level, there is no general epistemology that identifies the relevant values inherent in the right to privacy. Rather, the concept involves the protection of multiple interests and, as a consequence, a broad and disparate grouping of subject matters.[330] Thus, whereas the interests and content of rights and freedoms, such as the freedom of

[326] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 25.

[327] Diggelmann and Cleis (2014), p. 441.

[328] Brüggemeier (2010), p. 18.

[329] Brandeis and Warren (1890), p. 193.

[330] As a result of the flexible and broad nature of the right to privacy, various scholars have developed theoretical typologies to elucidate its content, affirming its characteristic as an umbrella

expression, are arguably easily identifiable,[331] the substance of the right to privacy is particularly flexible in international human rights law, continuously modified in accordance with social development.

Nevertheless, a broad typology demarcates privacy as being (1) associated with things we perceive we own: objects, homes, reputation and body parts, which should be inaccessible to others, unless a person consents; and (2) private choice, that is, decisional privacy to form the most important decisions concerning the intimate sphere.[332] This typology is similar to the approach by the ECtHR, which in case law has affirmed two broad categories of privacy—secrecy[333] and personal choice[334]—also found in other international fora.[335] For example, the UN HRC has stated that the right to privacy, more than solely guaranteeing a domain of seclusion, ensures 'a sphere of a person's life in which he or she can freely express his or her identity, be it by entering into relationships with others or alone'.[336] Similarly, the ECtHR has recognised the inclusion of several aspects relevant to forming a personal identity and the establishment and development of relationships with other human beings.[337] Given the range of protected values within the right to privacy, and concurrent obligations of non-interference and protection *vis-à-vis* such interests, it is not uncommon that conflicts arise between, for example, the right to anonymity and protection against harmful acts violating individual autonomy, to be discussed further in other chapters.

As viewed, individual autonomy is a central aspect of the right to privacy, which comprises physical integrity and a degree of control over personal information. The right to privacy is accordingly protected primarily on the basis of its instrumental importance to the individual, pertaining to a range of values such as the development of personality, autonomy and dignity.[338] Nevertheless, several authors also emphasise the social value of privacy.[339] Privacy enhances the capabilities of individuals to be reflective and develop new ideas, which in turn enriches social communication

concept. *See*, for example, Levmore and Nussbaum (2010), p. 10; Allen (1987), p. 464; Hughes (2012), p. 808; Diggelmann and Cleis (2014), p. 442.

[331] Solove (2006), p. 480.

[332] Allen (1987), p. 464.

[333] *Z v Finland* (1997) 25 EHRR 371 (a person's HIV-status).

[334] *Dudgeon v the United Kingdom* (ECtHR) (sexual orientation); *M. C. v Bulgaria* (2005) 40 EHRR 20 (sexual autonomy).

[335] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013) OEA/Ser.L./V/II. CIDH/RELE/INF.11/13, para. 131 and, more specifically, *Case of Artavia Murillo* et al. *v Costa Rica* (IACtHR) (reproductive freedom); *Case of Guzmán Albarracín* et al. *v Ecuador* (merits, reparations and costs) IACtHR Series C No. 405 (24 June 2020) (sexual violence).

[336] *Coeriel* et al. *v the Netherlands*, Communication No. 453/1991, UNHRC, UN Doc. CCPR/C/52/D/453/1991 (9 December 1994), para. 10.2.

[337] *X v Iceland* App no 6825/74 (Commission Decision, 18 May 1976).

[338] Allen (1999), p. 738; Hughes (2012), p. 809; Rappaport (2001), p. 443.

[339] Hughes (2012), p. 809; Lemmens (2003), p. 383.

and deliberative democracy.[340] By allowing individuals to develop their identity, relationships and intimacy are facilitated, as the person becomes better equipped to develop social connections.[341] Similarly, whereas privacy intrusions tend to focus on individual harm, the broader effect on society may also require state restraint. Certain acts, such as surveillance by state authorities, may not harm individuals directly while having a chilling effect on behaviour, for example, by altering the willingness to engage in political and social debate.[342] It may lead to self-censorship and the inhibition of creativity, thus affecting the greater social good. For example, removing anonymity online may lead to such consequences. As the freedom of expression is generally viewed as generating social benefits, it tends to trump the right to privacy in conflicts between the two rights. For example, if the right to privacy were solely viewed as a means of protecting private information, in instances where publication is in the interest of the public, the protection of privacy would more likely be overriden.[343] Underscoring the social and democratic values also of the right to privacy thus equalises the predisposed preeminence of the freedom of expression.

The right to privacy comprises protection against *invasions* of the private sphere, that is, not to receive information. This includes spam, under certain circumstances.[344] Meanwhile, protection of *secrecy* includes a right to anonymity, often in conjunction with the freedom of expression,[345] restrictions on collecting and processing private information as well as protection against disclosure. In terms of collecting and processing information, protection extends to the content of correspondence, including email and other online communication.[346] This primarily pertains to state and intermediary interference. Although it may be argued that the exchange of personal information is voluntarily surrendered in return for access to digital goods and services of ISPs/websites operators, this would presume that individuals can forego the use of electronic means of communication.[347] However, Internet access, whether construed as an independent human right or an aspect of existing human rights, is not solely a matter of individual choice. Non-interference also involves private persons or institutions, including the mass media. The confidentiality of communication on social media from surveillance by other individuals

[340] Schwartz (1999), p. 1658; Blume (2010), p. 158.

[341] Hughes (2012), p. 823.

[342] UNHRC, 'The Right to Privacy in the Digital Age: Report of the Office of the United Nations High Commissioner for Human Rights' (30 June 2014) UN Doc. A/HRC/27/37, para. 20; Solove (2006), p. 487.

[343] Hughes (2012), p. 827.

[344] *Muscio v Italy* App no 31358/03 (ECtHR, 13 November 2007).

[345] Solove (2006), p. 513.

[346] *Barbulescu v Romania* App no 61496/08 (ECtHR, 5 September 2017); *Copland v the United Kingdom* App no 62617/00 (ECtHR, 3 April 2007).

[347] Ronen (2015), p. 73.

has, for example, been affirmed by the ECtHR.[348] The publication, collection and processing of private information may also violate EU law. The evolving area of data protection derives from the right to privacy but diverges in that it applies to both the public and private sector and it is technologically specific in its regulation.[349] This specificity is evidenced by the EU Charter on Fundamental Rights, which contains two separate articles on data protection and the right to privacy.[350]

Furthermore, a person's image, be it photographs or videos, is also protected, whether involving taking or distributing images without consent. [351] This is particularly relevant in instances of image-based sexual abuse. While early case law of the ECtHR did not include protection of photographs taken in public places, this has given way to a more nuanced approach.[352] In the case of *Reklos and Davourlis v Greece*, the ECtHR held that:

> Whilst in most cases the right to control such use involves the possibility for an individual to refuse publication of his or her image, it also covers the individual's right to object to the recording, conservation and reproduction of the image by another person. As a person's image is one of the characteristics attached to his or her personality, its effective protection presupposes, in principle. . .obtaining the consent of the person concerned at the time the picture is taken and not simply if and when it is published. Otherwise an essential attribute of personality would be retained in the hands of a third party and the person concerned would have no control over any subsequent use of the image.[353]

The right to privacy also includes protection of a person's reputation. Reputation is considered an important aspect of self-identity and the ability to engage in public life and is specifically protected in the ICCPR, the UDHR and the American Convention on Human Rights (ACHR).[354] It was rejected as an aspect of the right to privacy in early case law of the ECtHR, albeit included as a legitimate aim for restrictions in Article 10 (2).[355] However, the Court in *Pfeifer v Austria* in 2007 held 'that a

[348] *Buturuga v Romania* (ECtHR) (domestic violence).

[349] *See*, for example, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *See*, generally, Blume (2010), p. 154.

[350] Art. 7 and Art. 8 of the Charter of Fundamental Rights of the European Union (the 'Charter'), 2010 O.J. (C83) 389.

[351] *Von Hannover v Germany* (ECtHR), paras. 108–113; *Sciacca v Italy* (2006) 43 EHRR 400, para. 29.

[352] *Friedl v Austria* (ECtHR), para. 49.

[353] *Reklos and Davourlis v Greece* App no 1234/05 (ECtHR, 15 January 2009), para. 40.

[354] Art. 17 of the ICCPR; Art. 12 of the Universal Declaration of Human Rights, Adopted and proclaimed by General Assembly resolution 217 A (III) of 10 December 1948; Art. 5 of the American Declaration of the Rights and Duties of Man (1948) O.A.S. Res. XXX, adopted by the Ninth International Conference of American States (1948), reprinted in Basic Documents Pertaining to Human Rights in the Inter-American System, OEA/Ser.L.V/II.82 doc.6 rev.1 at 17 (1992); Art. 11 of the American Convention on Human Rights (1969) O.A.S.T.S. 36, 1144 UNTS 123, entered into force 18 July 1978.

[355] *Lingens v Austria* (ECtHR).

person's reputation, even if that person is criticised in the context of a public debate, forms part of his or her personal identity and psychological integrity and therefore also falls within the scope of his or her "private life"'.[356] This includes protection against disclosure and defamation. Disclosure, for example, pertains to the publication without consent of the name and address of a person, personal photographs and information, such as a person's medical history or sexual orientation. Defamation concerns untruthful statements that harms the reputation of an individual. Nevertheless, defamation has mainly been addressed in relation to the freedom of expression by human rights law bodies, as a legitimate aim for state intervention.[357] It should in this regard be noted that disclosure of personal information and defamation commonly involve conflicts between the freedom of expression and the right to privacy, as the former includes a right to publish information and the latter a right to non-disclosure.

As for privacy as decision-making, regional human rights law courts and UN treaty bodies have affirmed that sexual and reproductive autonomy is an aspect of the right to privacy, in that it involves values such as decisional autonomy, physical integrity and human dignity.[358] While severe forms of sexual violence, such as rape, in certain instances also violate the prohibition on torture, inhuman or degrading treatment, all forms of sexual violence involve breaches of the right to privacy and thus engage positive obligations to protect.[359] Thus, gender-based online offences such as various forms of sexual violence, sexual harassment, defamation and the non-consensual distribution/receipt of intimate or other personal information, all transgress the right to privacy.

Many of the activities potentially in contravention of the right to privacy are not inherently harmful, but solely in cases of non-consent, be it in relation to data collection, disclosure of personal information or sexual activities.[360] Consent forms the core of such privacy related values as individual decision-making and control of access to personal information. This allows for a subjective approach to privacy, in that the degree of access and control is determined by each person. For example, the line between illegal and legal action in cases of so-called revenge pornography is non-consent, not the morality of the public display of nudity.[361] Whereas a consent-based standard is aligned with a focus on individual autonomy, challenges arise in assessing consent in relation to gender-based offences, whether IRL or on the Internet.

---

[356] *Pfeifer v Austria* App no 12556/03 (ECtHR, 15 November 2007), para. 35.

[357] For example, *Lingens v Austria* (ECtHR). *See* Sect. 4.3.4 for an overview.

[358] *X and Y v the Netherlands* (1986) 8 EHRR 235; *Case of Guzmán Albarracín* et al. *v Ecuador* (IACtHR); UN HRC, 'General Comment No. 28: Article 3 (The Equality of Rights Between Men and Women)', para. 20; *L.N.P. v Argentina*, Communication No. 1610/2007, UNHRC, UN Doc. CCPR/C/102/D/1610/2007 (16 August 2011), para. 13.7.

[359] *See* Sect. 4.2 on sexual violence.

[360] Gilden (2016), p. 484.

[361] ibid., p. 443.

In terms of state duties, these encompass both negative and positive obligations. The structure of assessing state intervention of the right to privacy is similar as in relation to the freedom of expression, as both are qualified rights in major international human rights law treaties.[362] In line with the development *vis-à-vis* the freedom of expression, the prominent attention given to data protection issues entails that much focus on the right to privacy on the Internet has been placed on negative obligations, with risks primarily viewed as that of surveillance, interception and the processing of communication by the state.[363] Such obligations affect available mechanisms of investigating and preventing offences and are thus relevant to this topic. There has been less discussion on positive obligations to *ensure* privacy through restrictions of certain types of behaviour online, particularly relevant in relation to gender-based online offences. Nevertheless, as will be made clear in Chap. 4, obligations to protect on the Internet may arise in relation to several violations discussed in the book. For example, obligations for states to protect individuals against sexual violence have been affirmed by regional human rights law courts and UN treaty bodies alike under the provision of the right to privacy, focusing on the right to sexual autonomy and physical and mental integrity.[364]

Meanwhile, dignitary harms such as defamation and the non-disclosure of private information do not generally engage obligations to protect, albeit under limited circumstances.[365] Rather, states are *allowed* to regulate, if they meet the requirements for restrictions on the freedom of expression. Accordingly, the categorisation of transgressions of privacy and the perceived harm affect the content of state obligations. For example, if instances of non-consensual disclosure of intimate photographs are construed as violations of sexual autonomy, positive obligations for states ensue.

[362] In order for a state to validly restrict the right to privacy, the basis must be provided in domestic law; the interference must be based on a legitimate aim; necessary in a democratic society and the means proportionate to the aim. Whereas the ICCPR does not enumerate legitimate aims, this is provided, for example, in Art. 8 of the ECHR. *See* UN HRC, 'General Comment No. 16: Article 17 (Right to Privacy), The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation' (8 April 1988) UN Doc. HRI/GEN/1/Rev.9 (Vol. I) (aims of the Convention).

[363] *See*, for example, UNHRC, 'The Right to Privacy in the Digital Age: Report of the Office of the United Nations High Commissioner for Human Rights' (30 June 2014). Although an individual may not have a reasonable expectation of privacy in public spaces, for example, *vis-à-vis* closed circuit TV:s recording public streets, it may be reasonable in relation to the processing of such information. The recording of data and the systematic or permanent nature of the record may violate the right to privacy. *See Peck v the United Kingdom* (2003) 36 EHRR 41, para. 59; *P.G. and J.H. v the United Kingdom* (2008) 46 EHRR 51, para. 57.

[364] *M. C. v Bulgaria* (ECtHR) (rape); *Rosendo Cantú* et al. *v Mexico* (IACtHR) (rape).

[365] There are certain exceptions, such as Art. 17 (2) of the ICCPR and Art. 11 (3) of the ACHR.

3.3.3.2 Privacy on the Internet

3.3.3.2.1 Introduction

The need for privacy does not arise in solitude but is socially created.[366] As privacy occurs in the context of social life, its content and values cannot be determined void of considerations of social development and practices.[367] For example, protection against the exposure and disclosure of private information is defined by social and cultural norms on what types of information are reasonable to protect as a safeguard of human dignity.[368] Reputation is a reflection of how a person is viewed socially and what information is considered harmful to reputation is thus both subjectively and objectively determined. Accordingly, the harm partly arises through social sanctions and condemnation. Such a relative approach has also been applied in international human rights law.[369] Due to the broad reliance on social influences informing the content of the right, privacy needs to be theorised contextually, bearing in mind the forum involved, such as the Internet. Changes in the social environment—such as new technology—require that the scope of the right must necessarily evolve.[370] As noted above, with technological innovation—evident also with the birth of the camera and telephone—individual expectations and social norms change. For example, caller ID was initially considered a privacy violation and is now generally viewed as privacy-enhancing technology.[371] Similarly, although there was an initial presumption that the Internet would strengthen privacy protection, since individuals were able to communicate anonymously, new technology has simultaneously generated elevated risks of intrusions upon privacy and reputation.[372]

Whereas it is clear that the right to privacy applies to the Internet, the question thus arises in which spaces on the Internet it is applicable and how the context affects its interpretation. Internet architecture and cyber norms both enhance and suppress certain values associated with the right to privacy, which may have an impact on

[366] Moore (1984), p. 73.

[367] Chang (2015), p. 147.

[368] Solove (2006), p. 534.

[369] UN HRC, 'General Comment No. 16: Article 17 (Right to Privacy), The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation' (8 April 1988), para. 7.

[370] Panichas (2014), p. 145; Hughes (2012), p. 806; Tene and Polonetsky (2014), p. 76; Blume (2010), p. 152. As noted by the CoE Committee of Ministers, it is expected that the interpretation of the right to privacy will evolve in line with developments in technology which may bring other forms of digital communication within the sphere of Art. 8 protection. *See* CoE, 'Recommendation CM/Rec(2014)6 of the Committee of Ministers to Member States on a Guide to Human Rights for Internet Users (Adopted by the Committee of Ministers on 16 April 2014 at the 1197th meeting of the Ministers' Deputies), Explanatory Memorandum', para. 80.

[371] Tene and Polonetsky (2014), p. 73.

[372] Solove (2010), p. 17; *Delfi v Estonia* (ECtHR), para. 110.

certain legal concepts and assessments, particularly in relation to the protection of reputation and sexual autonomy.

3.3.3.2.2 Spaces

From a broad perspective, the renegotiated boundaries of the right to privacy are affected by the delineation of public/private spheres. The social, cultural and political delineations of public and private spheres have shifted during the course of history.[373] The division is not natural but an expression of power.[374] It originates from social contract theories developed by philosophers such as John Locke and Jeremy Bentham. The theoretical foundation is thus liberalism, which distinguishes between public and private spheres of state influence, infused in domestic and international law, for example, codifying the right to privacy.[375] The public/private divide has traditionally been approached from a *spatial* standpoint, as a division between public areas and the home.[376] The scope of the sphere considered unsuitable for state regulation has traditionally involved marriage, family, reproduction and sexuality, associated with the home and historically considered the domain of women. In contrast, the public sphere of economic and political power has been denoted as predominantly male.[377] Whereas the spatial aspect of spheres—that is, the geographical and temporal distance between them—has decreased, which is evident in social media, surveillance and the work/home convergence, the current delineation still associates work with the "public" sphere and the home and personal life with the "private" sphere.[378]

The private sphere has, in fact, in many respects been redefined as an area in the life of each individual.[379] International human rights law bodies are increasingly approaching the private sphere from the viewpoint of subject matter, that is, the level of intimacy involved. Individual autonomy *vis-à-vis* others lies at the centre, with the right to privacy providing a degree of control through the creation of personal barriers.[380] Such sociological changes have thus had an impact on the content of the right to privacy. Nevertheless, in terms of assessing the scope of privacy, the delineation between public/private spheres of society is still relevant. In general,

[373] Roth (1999), p. 45.

[374] Koops and Galic (2017), p. 20.

[375] Roth (1999), p. 46.

[376] Public spaces are open to all, while access to private spaces is restricted and encompasses the human body and spaces personalised by people who inhabit them, such as houses, cars and computers. The basis for the spatial approach is the level of control over access to the space. *See* Koops and Galic (2017), p. 29; Ford (2011), p. 551.

[377] Allen (1999), p. 724.

[378] Ford (2011), pp. 554–555.

[379] Roth (1999), p. 45.

[380] Blume (2010), p. 153.

individuals still have a more limited right to privacy in public. At the same time, it is increasingly recognised that these spheres overlap. Private acts have public consequences and public regulations affect privacy interests.[381] For example, the ECtHR has held that the right is not strictly limited to the private sphere.[382] Accordingly, 'it would be too restrictive to limit the notion [of private life] to an "inner circle" in which the individual may live his own personal life as he chooses and to exclude therefrom entirely the outside world not encompassed within that circle'.[383]

According to the ECtHR, the right extends to the person's "reasonable expectation" of privacy in public spaces, which has been applied in relation to the Internet.[384] A "reasonable expectation" involves not only a subjective expectation of privacy, but an objectively reasonable one in light of existing social practices and values.[385] The expectation is a 'significant though not necessarily conclusive factor'.[386] For example, the ECtHR case of *Barbulescu v Romania* involved the monitoring of private instant messaging online between an employee and his fiancé and brother at his workplace.[387] This was deemed a violation of his private life, as the employee had not received adequate prior notice of the fact that his communication might be monitored. A reasonable expectation of privacy was thus affirmed in this public sphere. The Court categorised instant messaging as an aspect of 'private social life'.[388] However, the degree of privacy was considered different in public, as opposed to private places, with more extensive protection in the latter sphere.[389]

A general issue concerning the Internet is thus its place in the public/private dichotomy, which has an impact on the content of state obligations. As noted above, the Internet has characteristics of both spheres. In view of the increasing transfer of public functions online and its contribution to democracy, the Internet is generally considered a public sphere. Simultaneously, pockets of private spheres exist, for example, in the form of private emails or instant messaging. Such spaces are not public in the sense of being open to everyone. Privacy is also made public to a greater extent with individuals sharing personal information willingly or without consent on public social media, the Internet thus providing private spaces for speech

[381] Rappaport (2001), p. 449.

[382] *Brüggemann and Scheuten v Germany* App no 6959/75 (Commission Decision, 19 May 1976), para. 57.

[383] *Niemietz v Germany* (1993) 16 EHRR 97, para. 29.

[384] *P.G. and J.H. v the United Kingdom* (ECtHR), para. 57 (reasonable expectation); *Barbulescu v Romania* (ECtHR) (the Internet).

[385] Schauer (1998), p. 562.

[386] *Köpke v Germany* App no 420/07 (Commission Decision, 5 October 2010).

[387] *Barbulescu v Romania* (ECtHR).

[388] ibid., para. 74.

[389] Solove (2006), p. 496.

on a platform that, partly, serves a public purpose.[390] Furthermore, content frequently travels between private and public online spaces.[391]

Given this complexity, new theoretical models of categorisation are increasingly proposed *vis-à-vis* the right to privacy. Arguably, the distinction between the private sphere (as in the home and relationships) and the public sphere (as in the professional world), is no longer applicable to social life, which is characterised by a significant overlap and interaction. Several scholars have thus reconceptualised the public/private distinction as instead involving the intimate/social[392] or visibility/collectivity.[393] For example, it has been suggested that the Internet may be compartmentalised in such dichotomies depending on the audience. Private emails and chats would be treated differently than public blogs, newspapers and government information. As argued by William Mitchell, '[m]any of the places in cyberspace are public, like streets and squares; access to them is uncontrolled. Others are private, like mailboxes and houses, and you can enter only if you have the key or can demonstrate that you belong'.[394] Alternatively, the areas of the Internet may be understood as a continuum rather than a dichotomy, with the public and private worlds as anchors at either end.[395] For example, social media may be construed as publicised privacy rather than involving a public sphere.[396] The public/private divide is thus fluid in relation to the Internet and the particular pocket of online communication and publication of material is relevant when assessing the scope of privacy and the content of state obligations, evident also in the approach by human rights law bodies.[397]

3.3.3.2.3 Non-Disclosure

As mentioned, the protection of reputation encompasses regulation of the non-consensual disclosure of private information. However, the Internet affects the ability of users to control content already published, the right to have information removed and the level of harm to reputation.

In relation to the control of information—linked to the public/private distinction—the question arises whether the same demands for secrecy can be met

[390] Gagliardone et al. for UNESCO (2015), p. 8.

[391] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), fn. 109.

[392] For example, Andrea Slane suggests that the public sphere should be construed as a place where citizens can debate, complain, be confronted with new ideas, and the space of the home associated with the types of interactions that frequently occur in this location, that is, intimate communications with heightened levels of personal control. *See* Slane (2005).

[393] Weintraub (1997), p. 4.

[394] Mitchell (1995), p. 23.

[395] Ford (2011), p. 550.

[396] ibid., p. 560.

[397] *See* Chap. 4.

on a medium premised on sharing private information. Arguably, new theoretical thinking in relation to intimacy and secrecy interests in the Information Age is warranted.[398] This may either entail more rigorous or lower levels of protection of privacy, in addition to new forms of safeguards. The secrecy aspect of privacy considers information that is divulged as no longer private. As mentioned above, the offence of unlawful disclosure of private information entails that personal information becomes public. However, Daniel J. Solove argues that secrecy is given excessive prominence in theory and law as an aspect of the right to privacy.[399] The presumption that privacy is lost once information is disclosed does not fully translate to the Internet. Since much of our lives is now conducted in public fora, including intimate content which may be considered private from the viewpoint of subject matter, it is important that the right encompasses the ability to control the flow of information already divulged.[400] For example, the offence of "unlawful disclosure of information" should be adapted to encompass dissemination beyond a specific network of information flow. The harm of disclosure does not solely occur when complete secrecy is lost, but also when information is spread beyond expected boundaries. This is applicable to, for example, intimate photographs taken consensually but distributed without consent. Tentatively, such an approach is developing in international human rights law. For example, the ECtHR has held that a certain degree of protection of privacy may extend to already published material in instances of re-publication, to protect individual dignity.[401]

Considering reasonable expectations of the flow of information is beneficial in this regard.[402] Due to the close association between the right to privacy and social developments, several scholars have argued in favour of the application of discourses beyond law and philosophy when delineating the scope of privacy.[403] For example, Lior Strahilevitz argues that in cases involving limited disclosure of privacy, courts would benefit from considering empirical studies in sociology and psychology on the probability of information disclosed to one member of a particular network being further distributed to others.[404] This informs the question of whether the personal information would have been disseminated had the disclosure through the defendant not occurred. Arguably, information spreads in predictable ways. This considers the role of social norms in constraining or facilitating dissemination. It is common to share personal information, even embarrassing facts, with a close group

[398] Levmore and Nussbaum (2010), p. 11. *See* also Slane (2005), p. 258.

[399] Solove (2006), p. 537.

[400] Solove (2010), p. 20.

[401] *Aleksey Ovchinnikov v Russia* App no 24061/04 (ECtHR, 16 December 2010), para. 50.

[402] Solove (2006), p. 532.

[403] Hughes (2012), p. 806.

[404] This relates to both structural and cultural factors, including the prevalence and strength of social ties and the willingness to disclose facts to certain groups. Arguably, in the context of domestic courts in the US, a consistent methodology is rarely applied when examining reasonable expectations of privacy in limited privacy cases, that is, when shared with a few individuals. *See* Strahilevitz (2005), p. 919.

of people and having the expectation that it will not be shared.[405] Nevertheless, differences between offline and online social networks must be acknowledged. For example, the latter networks frequently involve larger groups with less intimacy[406] and social media involves a "context collapse" where individuals may find it difficult to compartmentalise their self-representation, as these often involve disparate social groups, such as friends, family and colleagues etc.[407] This may affect social norms and, in turn, legal assessments of, for example, reasonable expectations of privacy.

Meanwhile, Kirsty Hughes advances an approach to privacy that considers physical, behavioural and normative barriers that individuals choose in social interactions to preserve private space.[408] Accordingly, privacy does not revolve around the inclusion or exclusion of others but rather the possibility to control communication.[409] Generally, defining the scope of the right involves the objective identification of privacy-related interests. However, as the need for privacy in the main is a subjective experience, difficulties ensue. This model considers reasonable expectations of privacy in light of (1) whether there was an objectively recognised social norm that privacy should be respected and, if not; (2) consideration of the steps taken by the applicant, that is, physical/behavioural barriers to determine whether there was a reasonable expectation of privacy.[410] Technology must thus in turn be considered not only in relation to the development of social norms having an impact on reasonable expectations of privacy, but also in view of options for users to manage privacy. However, even with such an approach, the law must determine which invasions are sufficiently harmful to warrant regulation.

In contrast, it has been argued that expectations of privacy should be *reduced* in the Information Age and that social norms on reputation will adapt and allow for the presentation of more flawed and humanised versions of individuals. The increased publication of personal information may influence social norms to consider behaviour less embarrassing, such as in relation to sexual norms.[411] For example, the founder of Facebook, Mark Zuckerberg, has argued—when modifying privacy settings—that users no longer value privacy to the same extent.[412] The medium may thus have an impact on the development of social norms by reducing expectations of secrecy, with individuals sharing more private information voluntarily.

---

[405] Strahilevitz (2005), p. 925.

[406] McNealy (2012), p. 154.

[407] Tene and Polonetsky (2014), p. 91. There is also an overlap between personal and business spheres, as many employees access the Internet at work, for both personal and work-oriented purposes. *See* Yanisky-Ravid (2014), p. 57.

[408] Hughes (2012).

[409] ibid., p. 809.

[410] ibid., p. 814.

[411] Chander (2010), p. 135.

[412] The Guardian, 'Privacy no longer a social norm, says Facebook founder' (11 January 2010), <https://www.theguardian.com/technology/2010/jan/11/facebook-privacy> Accessed 14 March 2022.

However, as noted previously, shaming is reshaped on the Internet in a manner that increases its negative effects. This includes the absence of social norms restricting certain types of behaviour in face-to-face communications, the permanent recording of shameful behaviour which strips away context, and technological conduct—such as flaming and doxing—fueling moblike behaviour. This limits the possibilities for individuals to reinvent themselves and has a profound effect on the freedom of individuals to experiment, grow and change.[413] The ECtHR has also noted the increased risk of dignitary harm, including harm to reputation, with the widespread use of the Internet. [414] However, as noted above, the prevalence of virulent content online, including defamation, has simultaneously been considered to reduce the level of harm. Meanwhile, the forum in which the information was published has been deemed relevant by the ECtHR in assessing the harm of defamation or disclosure of information, including the size of the audience, for example, whether published on a well-visited or obscure blog.[415]

This threat of harassment and damage to reputation may cause individuals to refrain from activities that may lead to future embarrassment. Behaviour is thus modified in accordance with technology.[416] It may lead to either excessive caution—restricting decisional autonomy in the present—or reckless fearlessness—constraining choice in the future. The awareness that engagement in a position of authority in the public sphere may lead to the divulging of private information adversely affects decisional autonomy. In line with Michel Foucault's theory on the Panopticon prison, the inhibiting effect of surveillance functions as social coercion.[417] This is as such not constrained by the state but the social environment, limiting access to, for instance, employment, and affecting the social life and well-being of individuals.

The public disclosure of private information has particularly grave effects on young people. Adolescence is becoming increasingly public with young people less inhibited in divulging private information than previous generations.[418] The risks are also especially grave for women and girls, which has an impact on their engagement in the public sphere.[419] Since particularly women experience harassment, the non-consensual disclosure of private information and moblike bullying on the Internet, the impact on decisional autonomy is gendered.

In consideration of these arguments, the question arises whether it is reasonable for states to take measures to ensure or coerce privacy. Privacy online is to a large extent managed by users. Thus, the most efficient way to protect privacy is by dealing with it at its source, that is, to enable individuals to preserve their own

[413] Solove (2010), p. 17.

[414] *Payam Tamiz v the United Kingdom* (ECtHR), para. 80; *Delfi v Estonia* (ECtHR), para. 133.

[415] *Pihl v Sweden* (ECtHR), para. 37; *Payam Tamiz v the United Kingdom* (ECtHR), para. 88.

[416] Chander (2010), p. 127.

[417] Foucault (1995).

[418] Chander (2010), p. 127.

[419] ibid., pp. 126–127.

privacy. For example, social media users actively manage the level of dissemination of their private information. On Facebook, users may set their publicity settings to either public, private or restricted access. Such personal choices are beyond the purview of the law and the practical enforcement of privacy standards by states is limited. However, the limited protection of privacy on the Internet has led to demands for changes in technological design to increase privacy. This includes requests that social networking sites change their default settings.[420] For example, the Committee of Ministers of the CoE has framed the lack of privacy-friendly default settings as a human rights challenge *vis-à-vis* social media websites.[421] Technological means of ensuring consent by the object of photographs is also an option and is employed by certain website operators. [422] Nevertheless, Internet intermediaries are under no direct human rights law obligations apart from the duty to respect rights, affirmed in soft law documents. Obligations in this context thus extend to conducting human rights impact assessments of their services. In instances of transgressions of privacy, the regulation of individual accountability thus remains the main option for states.

3.3.3.2.4 Sexual Autonomy

While this will be explored further in Chap. 4, it has been affirmed by a range of regional human rights law courts and UN treaty bodies that the protection of sexual autonomy applies to the online sphere.[423] However, while the values to be protected are similar regardless of the forum, the Internet challenges traditional views of legal concepts, such as "sexual violence", "rape" and "non-consent", which may necessitate adaptation to the context of the Internet. Sexual autonomy may be violated in different ways online, for example, through the non-consensual disclosure of intimate images, receipt of unsolicited nude images and sextortion, typically in front of a web camera. An offline/online coherence thus requires ensuring that the protection of sexual autonomy does not solely extend to physical acts of sexual violence but also speech-based offences.

[420] Solove (2007), p. 200.

[421] CoE, 'Recommendation CM/Rec(2012)4 of the Committee of Ministers to Member States on the Protection of Human Rights with Regard to Social Networking Services' (Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies), para. 3.

[422] For example, Facebook. *See*, <https://www.facebook.com/safety/notwithoutmyconsent> Accessed 14 March 2022.

[423] *K.U. v Finland* (ECtHR). Meanwhile, the CEDAW Committee and the UN Special Rapporteur on Violence against Women have at a general level categorised sexual violence in this sphere as sex discrimination. *See* CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women'; UNHRC, '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).

In doing so, it is, however, necessary to consider the ways new technology has altered our approach to what are deemed appropriate forms of contact and visibility.[424] This informs practices of intimacy and relationships as the boundaries between acceptable conduct and intrusive communications are changing. For instance, although not widely researched, certain studies indicate that in the majority of cases of the non-consensual distribution of private sexual images, the victims have taken the photographs themselves.[425] While individuals are thus increasingly consenting to new technological sexual activities, such as "sexting", studies also demonstrate that mainly women are coerced into such acts.[426] There is a strong social pressure on women to share intimate images with their partners or others.[427] It is thus essential that assessments of sexual practices are not based on gendered presumptions of consent, where a victim's consent in one context is presumed to involve consent in a related context, an approach frequently adopted by law enforcement and reflected in social norms in instances of both image-based sexual abuse and sexual assault. [428] Liberalism and its emphasis on individual responsibility is an additional aspect, adding to the general reluctance to regulate non-consensual recordings or distributions of private sexual material in domestic laws.[429] The act of taking intimate photographs of oneself is often considered a moral lapse in judgment and—particularly for women—promiscuity.[430] A common approach is thus that victims have reduced their expectations of privacy by allowing the photographs or videos to be taken. Meanwhile, contextual assessments of, for example, non-consent in online interactions are challenging, particularly for intermediaries monitoring content but also in investigations by states.

3.3.3.3 Conclusion

The early cyberfeminist perspective presumed that cyberspace would provide desired privacy through anonymity and a space to reject expectations of female modesty. However, while both men and women are vulnerable to unwelcome privacy invasions, women do not enjoy privacy to the same extent nor the same

[424] Levy (2015).

[425] Cyber Civil Rights Initiative (CCRI), '2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration: A Summary Report', <https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf> Accessed 14 March 2022. *See* also Velez (2019).

[426] Laid et al. (2021) and Ross et al. (2019).

[427] Henry and Powell (2015), p. 107.

[428] Citron and Franks (2014), p. 348; Hill (2015), p. 123. For example, victim blaming is common in cases of date-rape, rape in relationships or in relation to the appearance and behavior of the woman.

[429] Arguably, '[v]ictims are disempowered through liberal narratives of personal responsibility and conservative values around sexuality, which all too often work together to reinforce a significant degree of victim blaming.' *See* Suzor et al. (2017), p. 1067.

[430] ibid., p. 1067.

type of privacy as men do IRL or online. It is clear that the content of the right to privacy, stemming from its theoretical foundation, has a male orientation associated with the public/private dichotomy. The gendered delineation of the private sphere, with more limited privacy in public, also creates particular challenges in relation to the Internet. Violations of international human rights law involve an act or omission by a state. However, the traditional approach to privacy has to a great extent excluded the privacy interests of women, by shielding the private sphere from state intrusions and thus private acts of violence, most commonly subjected to women. Simultaneously, women's autonomy in relation to intimate aspects of their lives, such as sexuality, has gone unrecognised. These issues have to a degree been resolved by the international human rights law regime, by affirming positive obligations for states to protect women against private forms of harm, in addition to recognising certain autonomy rights in matters intimately connected to the individual. Nevertheless, while such obligations are clear, for example, with regard to domestic violence and sexual violence, whether and how such concepts extend to online speech or acts, such as disclosure of intimate images and cyber sexual harassment, has been subject to limited analysis. Furthermore, positive obligations have only to a certain extent been affirmed in cases involving dignitary harm, such as reputation, prevalent on the Internet.

As noted above, the right to privacy in international human rights law protects a spectrum of values, such as psychological and physical integrity, decisional autonomy, human dignity and reputation. The majority of the gender-based online offences discussed in the book engage such aspects of the right to privacy. The offences mainly involve invasions of privacy (including disclosure, through the non-consensual distribution of intimate/personal information), exposure and distortion, which may harm reputation. Other intrusions on sexual autonomy are also addressed, such as physical acts of sexual violence and speech-based sexual harassment. However, not all transgressions are deemed equally harmful. For example, violations of sexual autonomy, such as rape, are considered particularly detrimental, while harm to a person's reputation is deemed less severe.

Furthermore, the level of harm is in certain instances assessed on a contextual basis, for example, when resulting from social harm, such as reputation. Whereas sexual violence is objectively harmful, harm to reputation, through *inter alia* defamation or disclosure of personal information, is determined *in casu* in view of the social setting. Similarly, reasonable expectations of privacy in public spheres are mainly assessed in relation to existing social practices and values. Where both harm and the scope of privacy are considered relative to social norms, care must be taken not to evaluate them in a manner detrimental to the recognition of online gender-based offences. For example, the ECtHR has accepted that coarser language is a social norm online. The current theoretical interpretation of privacy is also to a degree detrimental to women in relation to matters of secrecy on the Internet. As argued by Daniel J. Solove, the emphasis on secrecy must be reduced, in order to recognise invasions of privacy also where published information is non-consensually distributed beyond the intended audience. Privacy is less controlled on the Internet by the person it involves, and personal information is often

taken out of context, thus distorting its original meaning, while permanency entails that it is impossible to escape published information.[431] This must necessarily affect the recognition of harm, as well as appropriate solutions, such as "the right to be forgotten".

Although objective at its core, a contextual approach is also applied in relation to the harm of exposure and sexual intrusion, concerning the interpretation of non-consent and the assessment of the gravity of the harm. For example, norms on nudity have changed historically and still vary culturally, particularly in connection to gender. This may affect the perceived harm of exposure. The risk of affirming gender stereotypes must be balanced against recognising the particular harm associated with exposing female nudity. Social norms on intimate matters such as sexuality, health and relationships entail that women are commonly judged more severely, tainting their reputation to a higher degree.

Thus, while strides have been made to acknowledge a range of harmful conduct as violations of the right to privacy, challenges arise in applying standards and obligations to the Internet. This requires consideration of how the Internet affects certain forms of privacy-related offences. In view of the approach to reasonable expectations of privacy and the concept of harm *vis-a-vis* speech-based injury, there is a risk that the harm to integrity and autonomy online is not considered to rise to the level of a breach of the right to privacy.

3.3.4 Proportionality Assessments and Balancing in Conflicts of Rights

Rights interpretation in certain instances involves a review of the legitimacy of restrictions of rights, particularly pertaining to qualified rights, where the necessity of state measures is considered in relation to a specific aim. This includes a proportionality assessment. An additional aspect of rights interpretation is the balancing in conflicts of rights. As noted by Leto Carioulo, balancing constitutes substantive reasoning through which the scope of rights is determined, meaning that what is affirmed by the court in question is that a particular right can or cannot be exercised in a particular way.[432] In this manner, the resolution of conflicts of rights is an additional means of interpreting the content of provisions. Through balancing exercises, views on the level and nature of harm of speech or conduct may be inferred, also indicating the hierarchy of rights and interests in relation to the Internet. As proportionality assessments evaluate the legitimacy of state interferences—for example, the scope of criminal or civil laws or the regulation of intermediary liability—reasonable measures *vis-à-vis* online harm are also suggested.

[431] Crockett (2016), p. 174.

[432] Cariolou (2017), p. 173. *See* also Taramundi (2017), p. 122.

In terms of the methods employed in resolving conflicts, several approaches appear in international human rights law.[433] In certain instances, human rights law courts indicate their viewpoint as to which interest is supreme, both in the abstract and in the particular case, although this is not consistent. Such balancing may involve categorical prioritisation, which affirms abstract hierarchies of rights, for example, assigning absolute rights a higher status.[434] Regional human rights law courts have also adopted general legal principles for balancing interests in horizontal conflicts, where competing rights require equal respect, or in instances of intra-right conflicts.[435] Frequently, the balancing involves a consideration of whether a measure is the least restrictive means and proportionate to the aim.[436] In the case of the ECtHR, the scope of the margin of appreciation given to states affects the balancing exercise performed by the Court. States generally incur a wide margin of appreciation in conflicts of rights or interests.[437] More specific fair balancing tests have also been applied in relation to particular conflicts, such as between the freedom of expression and the right to reputation, as an aspect of the right to privacy.[438]

The Internet exacerbates tensions between rights and values, be it construed as a clash between liberty and equality, the freedom of expression and the right to privacy or intra-right conflicts.[439] As noted, gender-based offences on the Internet are mainly speech-based and frequently involve transgressions of autonomy, which is an aspect of the right to privacy. Their regulation thus often requires restrictions of the freedom of expression. Meanwhile, the architecture of the Internet is aligned with the value of the freedom of expression to individual autonomy, allowing users greater control over online communication and design, and more limited means of state restrictions of speech.[440] Self-governance realises the ideals of the liberal democratic state, that is, a political order based on the primacy of local norms and individual choice.[441]

---

[433] There are general theories on how to balance interests, for example, in relation to the freedom of expression, which includes "maximizationalist consequentialism" or "prioritarianism". The first concerns an optimum balance between interests of the speaker, audience, and third parties. The latter asserts that costs and benefits in relation to those worst off in society are supreme, even if there is a net benefit for unrestricted speech. *See* Brown (2015), p. 223; Brax (2016), p. 188. These theories are generally not applied by international human rights law institutions.

[434] *Gäfgen v Germany* (2010) 52 EHRR 45, para. 107: 'The philosophical basis underpinning the absolute nature of the right under Article 3 does not allow for any exceptions or justifying factors or balancing of interests, irrespective of the conduct of the person concerned and the nature of the offence at issue.'

[435] These exercises are often *ad hoc*-based, open-ended and, arguably, subjective. Calls for more structured reasoning have been made. *See*, for example, Smet (2010); Brems (2017), p. 248.

[436] Specifically in relation to limited rights: Art. 8–11 of the ECHR.

[437] *Evans v the United Kingdom* (2008) 46 EHRR 34, para. 77.

[438] *Axel Springer AG v Germany* (2012) 55 EHRR 218, paras. 89–95.

[439] Rona and Aarons (2016), p. 503.

[440] Balkin (2004), p. 52.

[441] Netanel (2000), p. 402.

This reinforces conflicts with rights advancing substantive equality, a principle which requires state intervention.

In such conflicts, the above sections indicate that balancing in many instances favour the non-regulation of this sphere. A similar approach is noticeable in the proportionality assessments of restrictions of rights affecting the Internet. For example, the ECtHR has in a range of cases involving competing interests assigned an *a priori* importance to the freedom of expression by emphasising its fundamental value to democracy.[442] Additionally, given the standing of the Internet as a public or quasi-public sphere, the protection of the freedom of expression is enhanced when restrictions may affect the operation of the Internet. As will be addressed further, this approach has, tentatively, been adopted in the case law of the ECtHR and by other international organisations.[443] According to the IACmHR, '. . .the original and special characteristics of the Internet should be taken into account before making any regulation that would affect its architecture or interaction with society'.[444] As such, the beneficial features of the Internet, such as direct democratic participation, must be borne in mind given the unprecedented possibilities for individuals to impart/seek information and its 'enormous capacity to serve as an effective platform for the fulfilment of other human rights'.[445] As such, '. . .it is crucial to evaluate all legitimacy conditions of the limitation of the right to freedom of expression based on these unique and special characteristics'.[446] Therefore, when assessing the proportionality of a restriction, it is necessary to consider the impact and cost not only from the viewpoint of the parties directly affected but also from the perspective of the operation of the Internet. A restriction may be minor from the standpoint of the individual but may have major consequences for the general functioning of the Internet. Each restriction must thus be evaluated from a 'systemic digital perspective'.[447] This has in turn affected the scope of state obligations to regulate online content and, implicitly, liability for intermediaries and media publishers.

Furthermore, states maintain a wider degree of discretion in regulating public as opposed to private spheres, given the premise of non-interference in the private

[442] *Sürek v Turkey (No.1)* App no 26682/95 (ECtHR, 8 July 1999), para. 61; *Lindon and Others v France* (2008) 46 EHRR 35, para. 46; *Axel Springer AG v Germany* (ECtHR), para. 90.

[443] For example, *Payam Tamiz v the United Kingdom* (ECtHR), para. 90; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. David Kaye' (30 March 2017) UN Doc. A/HRC/35/22, paras. 47–48

[444] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 12.

[445] ibid., para. 53.

[446] ibid., para. 53

[447] ibid., para. 53. *See* also 'Joint Declaration on Freedom of Expression and the Internet', signed by the UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and Special Rapporteur on Freedom of Expression and Access to Information and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information on 1 June 2011, principle 1 (b).

realm of individuals. The public sphere connotations of the Internet have been deemed to broaden the margin of appreciation for states in the regulation of Internet intermediaries, for the benefit of the freedom of expression.[448] Similarly, the dissenting opinion of several judges in the case *Mouvement Raëlien v Switzerland* argued that 'the Internet being a public forum par excellence, the State has a narrow margin of appreciation with regard to information disseminated through this medium', in terms of restricting online speech. [449] The nature and aim of the Internet thus result in a predisposed generous approach to the freedom of expression online.

This consideration of the value of the Internet affects the application of women's rights online at a general level. In relation to the balancing in conflicts of rights or values, the feminist legal critique *vis-à-vis* both domestic and international human rights law includes an opposition to the *a priori* assignment of certain values to rights, rather than consideration of the concrete effects in a particular case or context. The abstract affirmation of rights hierarchies and the values attached to rights may disadvantage women, as they commonly rest on male norms and ideologies and fail to consider the context of gender subordination.[450] The emphasis on the value of the Internet as being integral to democracy may entail that conflicts become prejudiced at an abstract level to the benefit of a generous approach to the freedom of expression and an unregulated Internet. Meanwhile, as noted previously, the online experiences of female victims are frequently undervalued, especially when involving speech-based harm, entailing a categorisation of a right lower in the hierarchy of rights. Harm emanating from speech—although causing psychological injury and negative effects on gender equality—is generally viewed as minor. With the public world clearly superior in the public/private hierarchy, harm suffered by women is more prone to be ignored or trivialised.

At the same time, the migration of the private into the public sphere increases the demands for more extensive protection and state involvement. The public value of the Internet entails that acts of violence cannot merely be categorised as private incidents, since they occur in a public domain of importance to democracy, enhancing obligations for states. As argued earlier, the significance of the Internet as a public sphere must accordingly produce positive obligations not only in terms of ensuring equal access and participation but also an environment free of harmful content and communication. Such an argument has, for instance, been made in EU law, in calls for regulating sexual harassment on the Internet.[451] In this regard,

[448] *Payam Tamiz v the United Kingdom* (ECtHR), para. 90.

[449] *Mouvement Raëlien Suisse v Switzerland* (ECtHR), Dissenting opinion of Judge Pinto de Albuquerque.

[450] Baines (2009), p. 31.

[451] European Parliament resolution of 11 September 2018 on measures to prevent and combat mobbing and sexual harassment at workplace, in public spaces, and political life in the EU (2018/2055(INI)).

balancing exercises must consider the impact of an unrestricted Internet on substantive gender equality.[452]

As will be viewed, beyond such value-based considerations, primarily the ECtHR but also other international sources frequently take a pragmatic approach to Internet regulation in proportionality assessments. For example, user anonymity has been considered a reason for extending liability to intermediaries and media publishers.[453] Meanwhile, the ECtHR has in assessments of the reasonableness of state regulation of intermediary liability taken note of available techniques for monitoring and moderating content, the ability of intermediaries to evaluate the illegality of particular types of content and the characteristics and business model of the online platform in question.[454] In this regard, the features of Internet architecture have been factors driving the legal assessment of liability.

3.3.5 Summary

Certain broad conclusions to the sections on harm, the scope of rights, proportionality assessments and balancing exercises can be drawn. The concept of harm, as applied at the domestic and international levels, minimises the harm of speech-based offences—thus the majority of online violations—and specifically its gendered consequences, be it in relation to individual women or women as a group. Meanwhile, although provisions on gender-based violence and gender stereotyping are applicable online, most online offences engage the freedom of expression and the right to privacy. The values, scope and hierarchy of these rights are, in certain respects, interpreted in a gendered manner. In turn, the Internet enforces certain values of the freedom of expression and the right to privacy while undermining others, in a way that exacerbates online harm and impedes its regulation. In proportionality assessments and balancing of rights, in many instances, the value of the freedom of expression prevails over other interests, especially online or when potentially affecting the operation of the Internet. With speech-based and gendered offences viewed as causing minor harm—and the broader impact on gender equality rarely addressed—regulation is thus often considered disproportionate. The difficulties in controlling harmful online speech, in view of Internet architecture, additionally hampers incentives to restrict speech. This is also a consideration in

[452] For example, structured balancing of interests. Stijn Smet proposes the consideration of the impact on each right by allowing the other to take preference, the involvement of other rights and the purpose of the exercise of the right. *See* Smet (2010). Meanwhile, "prioritarianism" involves considering equality as an aspect in measuring utility, that is, even if there is a net benefit for unrestricted speech, the fact that the costs are placed on those who are disadvantaged would tip the balance. *See* Brax (2016), p. 188.

[453] *Delfi v Estonia* (ECtHR).

[454] For example, *Delfi v Estonia* (ECtHR), para. 117 (large media publisher and clearly unlawful speech); *Pihl v Sweden* (ECtHR), para. 31 (limited readership).

proportionality assessments. These legal considerations and exercises must consequently be infused with a gender-sensitive approach, in order to fully acknowledge the gendered consequences of harm on the Internet.

3.4 Who Is Liable?

3.4.1 Introduction

In addition to the concept of harm and the scope of rights, significant obstacles of an ideological and practical nature arise in regulating liability for cyber offences, linked to the characteristics of international human rights law and Internet architecture. The following part will explore limitations to regulating liability in international human rights law at a general level, applied further in Chap. 4 on various online offences. The focus lies on intermediary liability, with positive obligations *vis-à-vis* individuals explored more in-depth in the subsequent chapter. It serves to highlight existing gaps in regulation and, to a more limited extent, potential means of resolving issues undermining protection.

As noted above, the public/private distinction exists as a theoretical framework in multiple philosophical disciplines, including geography, sociology, and law. In relation to international human rights law, the dichotomy arises in terms of the subjects of law and the approach to acceptable state involvement in the private sphere. With its foundation in liberal ideology—aiming to protect individuals against state interference—the focus lies on acts perpetrated in the public sphere by the state. This in turn affects which areas are subject to regulation, the scope of legitimate state interference and the content of obligations for states. These boundaries cause particular challenges in relation to online gender-based violations which are, regardless of context, commonly perpetrated by private individuals. While the development of positive state obligations to prevent violations between private individuals has been essential in this regard, the state remains the subject and these forms of obligations are not as far-reaching, with gaps remaining in relation to a range of interpersonal offences.[455] Furthermore, whereas state obligations to criminalise, investigate and prosecute such violations remain on the Internet, from the viewpoint of liability, anonymity creates a practical impediment to the effective investigation of interpersonal offences online. This will thus be addressed in the following.

Meanwhile, Internet intermediaries and media publishers provide the platforms through which offences are perpetrated and harmful material distributed. As

[455] Developed initially in the case of *Velasquez Rodriguez v Honduras* (merits) IACtHR Series C No 1 (29 July 1988), subsequently adopted by the ECtHR in *X and Y v the Netherlands*. Such obligations have also been adopted as legal standards by international and regional organisations, such as the UN and the CoE.

non-state actors, they are also not subjects of international human rights law. Although positive obligations to regulate also intermediaries are increasingly being developed, practical constraints arise in this regard, with Internet architecture affecting the capabilities of such entities to control content and conduct. The structure of international human rights law is thus not neutral in its effect, undermining the effective regulation of a range offences primarily experienced by women.

3.4.2 Individual Perpetrators and User Anonymity

The main challenge in ensuring the effective investigation and prosecution of individual perpetrators of online offences is user anonymity. Anonymity is the condition of avoiding identification.[456] Genuinely anonymous communication online is rare, as it requires the user to employ advanced technical measures.[457] There are various forms of anonymity, for example, being anonymous to the public but identifiable by a service provider through registration, or broader anonymity where users do not have to identify themselves but can be identified through information retained by ISPs, which may require an injunction.[458] Diverse means of ensuring secrecy are available, for instance, tools to mask IP-addresses, such as Virtual Private Networks (VPN), proxy services, anonymising networks and software, as well as peer-to-peer networks.[459] This concerns both speakers and readers.

A right to anonymity—affirmed by a range of international human rights law bodies—is considered an aspect of multiple human rights, including the right to privacy and the freedom of expression.[460] It also allows individuals to partake in such rights as voting and the freedom of association without fear of reprisals.[461] As an aspect of the right to privacy, the protection of anonymity and encryption are encompassed in the safeguarding of secrecy, as a means of reducing the risk of surveillance. Data protection, as a standalone norm or an aspect of the right to privacy, ensures protection against the collection and processing of such

[456] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015) UN Doc.A/HRC/29/32, para. 9.

[457] Gagliardone et al. for UNESCO (2015), p. 15.

[458] *Delfi v Estonia* (ECtHR), para. 148.

[459] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), paras. 7 and 9.

[460] Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights, 'Standards for a Free, Open and Inclusive Internet' (15 March 2017) OEA/Ser.L/V/II.149, para. 227; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015); *K.U. v Finland* (ECtHR); *Benedik v Slovenia* App no 62357/14 (ECtHR, 24 April 2018).

[461] Solove (2006), p. 513.

information.[462] As anonymity is an integral feature of the Internet, built into its architecture, it has also become a social norm on the Internet, which may have an impact on the assessment of reasonable expectations of privacy.

Anonymity as an aspect of the freedom of expression ensures the free exercise of public debate. For example, the Special Rapporteur on the Freedom of Expression of the IACmHR has recognised the right to freely express opinions and seek information 'without being forced to identify him or herself or reveal his or her beliefs and convictions or the sources he or she consults'.[463] Anonymous (or pseudonymous) expression has a long tradition in the written press, including by authors of books and critical reports, such as whistle-blowers.[464] From a democracy aspect, anonymity is important since valuable information or viewpoints may otherwise be chilled, hampering benefits to society and the audience, such as the search for truth.[465] As discussed previously, ICTs enhance the possibilities for women to participate in public fora and engage on political issues. This is especially the case for vulnerable or at risk groups such as journalists, civil society organisations, human rights activists and scholars, who may circumvent unlawful state-imposed censorship.[466] Women who live under a threat of violence may find safety in anonymity, establish virtual communities and participate in public debates.[467] Anonymity is also useful in enabling survivors of gender-based violence to seek help and access information, as well as to provide a space for women to express their views on issues constrained by gender stereotypes and social taboos IRL.[468] The UN Special Rapporteur on Violence against Women and the UN High Commissioner for Human Rights have thus emphasised the importance of anonymity in guaranteeing women's freedom of expression and privacy.[469] Accordingly, while friction often arises between the freedom of expression and the right to privacy in the context of the Internet, there

[462] As noted in Sect. 3.3.3, this is regulated through Art. 8 of the Charter on Fundamental Rights of the EU and GDPR, but has also been interpreted within the scope of the right to privacy, for example, by the ECtHR.

[463] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 132.

[464] Benedek and Kettemann (2013), p. 37.

[465] Levmore (2010), p. 60.

[466] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 12.

[467] UNHRC, 'Report of the Working Group on the Issue of Discrimination against Women in Law and in Practice' (19 April 2013), para. 48.

[468] UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017) UN Doc. A/HRC/35/9, para. 22.

[469] UNHRC, '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), para. 60; UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 18.

is not necessarily an incompatibility between these rights *vis-à-vis* anonymity, although intra-right conflicts occasionally arise. [470]

As a consequence, prohibitions of the individual use of encryption and anonymity disproportionately restrict the freedom of expression, since they deprive all online users the right to 'carve out private space for opinion and expression' without a specific claim of unlawful use.[471] For example, the ECtHR has noted that filter-bypassing technologies are content-neutral and may be used for both malevolent and legitimate purposes, with state suppression of information disproportionately restricting the freedom of expression.[472] In this regard, *blanket* removals of anonymity by the state have been considered disproportionate.[473] This encompasses requirements of real name registration for online activity, or the ban of anonymising tools, which in effect undermine the possibility of remaining anonymous.[474] Nevertheless, such restrictions do not extend to private companies, such as social media, which often condition access on real name registration. For example, more rigorous requirements of social media sites and dating apps to confirm the identity of users are increasingly employed as a safety precaution.[475]

While anonymity is integral in ensuring secrecy, its availability limits possibilities to protect individuals against harm that violates other aspects of privacy, such as sexual autonomy. As noted above, anonymity is a key technical feature in the consumption and distribution of child pornography, cyber harassment and cyberbullying.[476] Certain scholars have as a consequence argued in favour of reducing the possibilities of remaining anonymous on the Internet, under narrowly defined circumstances.[477] In international human rights law, the protection of individuals from the collection of personal information is balanced against the interests of states to prevent crime, as well as state obligations to criminalise, investigate and

[470] For example, with regard to the freedom of expression, theories on the marketplace of ideas and democracy mainly serve the interests of the audience, not the speaker. A right to anonymity for speakers may thus be opposed to audience rights to receive information. Knowing the identity of the speaker is arguably also important in order to assess the credibility of the opinions and to evaluate the worth of the information and ideas. *See* Barendt (2016), pp. 62, 66.

[471] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 40.

[472] *Engels v Russia* App no 61919/16 (ECtHR, 23 June 2020), paras. 29–30.

[473] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 40.

[474] ibid., para. 50.

[475] Powell and Henry (2017), p. 87.

[476] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 13.

[477] For example, Martha Nussbaum argues that as anonymity is the key problem of online offences, identification as a condition of posting should be required. *See* Nussbaum (2010), p. 85. Meanwhile, Solove proposes that a record of IP addresses should be kept, ensuring traceability in compelling cases, while still maintaining anonymity. *See* Solove (2007), pp. 146–147.

prosecute.[478] As anonymity is encompassed in two qualified rights, it may be restricted. Bearing in mind state obligations to protect individuals against certain forms of interpersonal harm, an appropriate measure in fulfilling the duty to investigate may thus be to require the identification of users. The identification of an offender on the Internet is primarily achieved through establishing a person's Internet Protocol (IP) address, information accessible to ISPs. Whether state obligations to investigate include obliging intermediaries to disclose information must thus be clarified. However, in relation to *which* offences anonymity may be removed and *how* anonymity can be lifted has not been extensively explored in international human rights law. Nevertheless, it is clear that it involves a balancing between opposing protected interests, where the nature of the offence is a consideration.[479]

The ECtHR has considered this issue in several cases involving sex-based offences against children. In *K. U. v Finland*, the ECtHR affirmed a right to anonymity as an aspect of the right to privacy, while noting its non-absolute nature.[480] The case concerned the placement of an advertisement by an anonymous person on an Internet dating site in the name of a minor, without his knowledge. The boy as a result received an email from a man who wished to meet him. The applicant's father requested that the police identify the person who had placed the advertisement in order to bring charges. The ISP did not, however, divulge the identity of the author of the advertisement as it considered itself bound by domestic law to respect the confidentiality of the information. The domestic court concurred that there was no explicit provision authorising the service provider to disclose identification data. Even though domestic law provided a right for the police to obtain such information in relation to certain offences, "malicious misrepresentation"—applicable in the case—was not included.[481]

The ECtHR affirmed that the development of telecommunications technologies in recent decades has led to the emergence of '. . .new types of crime and has also enabled the commission of traditional crimes by means of new technologies'.[482] It

[478] *See,* for example, *K.U. v Finland* (ECtHR). *See* also CoE, 'Explanatory Report on the Convention on Cybercrime' (ETS No. 185) 23 November 2001, para. 62. The CoE affirmed the principle of anonymity in its CoE Committee of Ministers, 'Declaration on Freedom of Communication on the Internet' (28 May 2003), Principle 7. However, the right to anonymity '. . .does not prevent states from taking measures to trace those responsible for criminal acts, in accordance with national law, the ECHR and other international agreements in the fields of justice.' The Budapest Convention does not criminalise the use of computer technology for purposes of anonymous communication. However, according to its Explanatory Report, '. . .Parties [to the Budapest Convention] may wish to criminalise certain abuses related to anonymous communications, such as where the packet header information is altered in order to conceal the identity of the perpetrator in committing a crime.' *See* Explanatory Report to the Convention on Cybercrime, ETS No. 185, para. 62.

[479] *Delfi v Estonia* (ECtHR), para. 149.

[480] *K.U. v Finland* (ECtHR).

[481] The possibility of bringing a lawsuit against the service provider existed, as providers were required to verify the identity of an author in cases of defamatory advertisements on their websites. However, this offence had become time-barred.

[482] *K.U. v Finland* (ECtHR), para. 22.

noted that, as a result, the CoE Committee of Ministers has adopted recommendations on criminal procedural law, obliging states to require that service providers who offer services to the public release information that identifies the user, when ordered by authorities.[483] The EU Directive 2006/24/EC on data retention in publicly available electronic services or public communication networks also obliges Member States to ensure that certain categories of data are retained for a period between 6 months and 2 years. It applies to data necessary to identify the user, rather than the content of electronic communications.[484] In reviewing the laws of Member States, the ECtHR concluded that the majority of domestic laws oblige telecommunication service providers to submit subscriber information in response to a request by investigative or judicial authorities.[485]

From this standpoint, the Court examined the content of positive obligations to protect individuals against interpersonal violence. Although states generally retain a broad margin of appreciation with respect to positive obligations, in relation to '. . .grave acts, where fundamental values and essential aspects of private life are at stake', the adoption of criminal law provisions are necessary in order to provide effective deterrence.[486] In this case, the domestic criminal law comprised an applicable offence. However, it was held that criminal law provisions have limited deterrent effect if there are no means with which to identify the offender and prosecute the individual.[487] Importantly, the Court noted that although it was possible to sue the service provider for damages, it was not sufficient. Rather, '[i]t is plain that both the public interest and the protection of the interests of victims of crimes committed against their physical or psychological well-being require the availability of a remedy enabling the actual offender to be identified and brought to justice. . .and the victim to obtain financial reparation from him'.[488] Measures to identify perpetrators of sexual abuse, particularly involving children, was thus considered an obligation, with such interests overriding the right to confidentiality.

The ECtHR has also considered the demand for the release of IP addresses by ISPs as legitimate in cases of child pornography, with state interference similarly

[483] CoE, 'Recommendation No. R (95 13 of the Committee of Ministers to Member States Concerning Problems of Criminal Procedural Law Connected with Information Technology' (Adopted by the Committee of Ministers on 11 September 1995 at the 543rd meeting of the Ministers' Deputies), para. 12. UN General Assembly Resolutions similarly provide that states should permit access to electronic data pertaining to particular criminal investigations. *See*, for example, UNGA, 'Resolution adopted by the General Assembly: Combating the Criminal Misuse of Information Technologies' (22 January 2001) UN Doc. A/RES/55/63.

[484] Art. 5 and Art. 6 of Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC (2006) OJ L 105/54.

[485] *K.U. v Finland* (ECtHR), para. 32.

[486] ibid., para. 43.

[487] ibid., para. 46.

[488] ibid., para. 47.

pursuing the aims of the "prevention of crime" and the "protection of the rights of others".[489] Nevertheless, rule of law principles must be respected, as made apparent in the case of *Trabajo Rueda v Spain* where the seizure and inspection of computer files by the police without prior judicial authorisation was considered disproportionate to the legitimate aim.[490] Similarly, in *Benedik v Slovenia*, a failure by Slovenian police to obtain a court order before requesting subscriber information in relation to a dynamic—that is, temporary—IP address from an ISP violated the applicant's right to privacy, more specifically the requirement of the interference being "in accordance with the law".[491] The applicant was suspected of sharing child sex abuse material through a peer-to-peer file-sharing network and even though the state thus had a legitimate aim—namely the protection of the rights and freedoms of others—it had not respected the rule of law. The ECtHR again took note of the expansive regulation on personal data protection and privacy of electronic communications within the EU and considered the concept of "personal data" to involve information relating either to identified or identifiable individuals.[492] From this standpoint, it considered 'whether the applicant, or any other individual using the Internet, had a reasonable expectation that his otherwise public online activity would remain anonymous'.[493] Interestingly, the assessment was performed 'independently from the legal or illegal character of the activity in question', as well as 'without any prejudice to the Convention's requirement that protection of vulnerable individuals must be provided by the member States'.[494] The Court concluded that Benedik did have a reasonable expectation of privacy in that, even though his IP address was visible to other users of the network, it could not be traced to a particular computer without obtaining information from the ISP, that is, a presumption of anonymity that was not undermined by the fact that he did not hide the IP address, as argued by the state.

A more nuanced approach was taken by the ECtHR in *Standard Verlagsgesellschaft MBH v Austria (No. 3),* in which it upheld the right of an online news portal to refuse to disclose data of anonymous users for allegedly defamatory statements.[495] Although publishing its own journalistic content, the company was categorised as a host provider by the domestic courts, in terms of the discussion board. Meanwhile, the ECtHR found a link between these two functions, with the comment section furthering the aim of open discussions on topics of public interest. In order the leave comments, users had to register on the news portal, including their name and email address. According to company standards, such data would be

[489] *Benedik v Slovenia* (ECtHR).

[490] *Trabajo Rueda v Spain* App no 32600/12 (ECtHR, 30 May 2017).

[491] *Benedik v Slovenia* (ECtHR).

[492] ibid., paras. 46–62.

[493] ibid., para. 98.

[494] ibid., para. 99.

[495] *Standard Verlagsgesellschaft MBH v Austria (No. 3)* App no 39378/15 (ECtHR, 7 December 2021).

disclosed if required by the law. In the case, a domestic court order obliged the publisher to divulge user data of anonymous commenters in conjunction with its articles, as the comments appeared to be defamatory, although no final decision on the lawfulness had been taken. Although the company removed the comments, it did not submit the data.

The ECtHR held that the domestic courts had failed to balance the interests of the plaintiffs on the one hand and the authors of the comments and the company on the other hand. Although the Court affirmed a right of access to justice for victims of defamation, domestic courts must consider certain factors when balancing interests. The function of anonymous speech in fostering the freedom of expression must be taken into account, as a means of avoiding reprisals or unwanted attention, not only in general but also in the particular context.[496] Furthermore, the Court considered that the nature of the comments was not decisive, appearing to weigh less as a factor when involving the removal of anonymity than when concerning secondary liability for intermediaries. Accordingly, the fact that the lawfulness of the comments had not been assessed by the domestic court was not crucial. At the same time, the Court noted that although offensive, the statements did not amount to hate speech or incitement to violence.[497] Rather, the comments could be characterised as political speech and were not clearly illegal. A significant factor was thus that the comments were part of a political debate, in connection to articles on political content, leading to a narrow margin of appreciation. The disclosure of personal data can have a chilling effect on discussions of a public interest. The same conclusion may thus not be reached where defamation concerns other subject matters and in other contexts. The Court also indicated that such balancing would not be relevant in instances of hate speech or other clearly unlawful speech.[498]

These cases thus indicate a balancing of interests where the nature of the offence, the identity of the victim and the operation of the Internet are considered. Arguably, where the autonomy interests involved in anonymity is low, with a limited impact on the core of the right, and the risk of abuse is high, states may require disclosure of data identifying users.[499] The removal of anonymity was deemed legitimate in cases involving sexual violence, given the severity of the harm, coupled with the vulnerability of victims, with states acquiring particularly extensive obligations to protect children against sexual offences.[500] Furthermore, *Standard Verlagsgesellschaft MBH v Austria (No. 3)* indicates that intermediaries or media publishers would be required to divulge the identity of alleged perpetrators in cases of hate speech—also

[496] ibid., para. 95.

[497] ibid., para. 89.

[498] ibid., para. 95.

[499] Lidsky and Cotter (2007), p. 1592.

[500] Meanwhile, the categorisation of women as a vulnerable group is controversial. Such an approach has, for example, been rejected by the ECtHR while supported by other international bodies. *See Valiuliene v Lithuania*, App no 33234/07 (ECtHR, 6 March 2013), para. 69. In contrast, for example, UN Office of the High Commissioner, <https://www.ohchr.org/EN/Issues/Health/Pages/GroupsInVulnerableSituations.aspx> Accessed 14 March 2022.

in view of the level of harm—or other cases of clearly unlawful speech. It is implied that this may encompass defamation in certain contexts. It thus appears that in the majority of the offences discussed in the book, states may either have the option or obligation to ascertain the identification of perpetrators. Nevertheless, as this does not turn solely on the nature of the offence, it will be determined on a case-by-case basis.

3.4.3 Liability of Internet Intermediaries and Media Publishers

3.4.3.1 Soft Law Obligations for Intermediaries and Self-Regulation

In view of the powers and capabilities of Internet intermediaries to shape the online environment and the practical difficulties of holding individual perpetrators accountable, the development of liability mechanisms for intermediaries and online media publishers is increasingly raised as a means of preventing harm. Additionally, given the technosocial effects of the Internet, the influence on user behaviour may most effectively be directed through code and intermediary control.[501]

Although international human rights law treaties regulate the relationship between states and individuals within their jurisdiction, a prominent feature of the Internet is the decentralisation of power. Rights, such as the freedom of expression, are ensured mainly by private entities, yielding extensive authority and control over a public sphere. Online gender-based offences thus transpire in an environment largely beyond state control. As emphasised by international organisations, Internet governance is a shared responsibility.[502] Internet intermediaries provide services that enable and mediate online communication, without generating or controlling content. They include ISPs, search engines, web hosting providers and social media platforms. As such, they serve various functions: ISPs connect a user's device to the Internet; web hosting providers make it possible for websites to be published; search engines allow individuals to search the World Wide Web and social networks facilitate interaction between Internet users.[503] These entities are thus not publishers and are separate from other media that create and disseminate original content, for instance, news websites that publish articles written by staff. They are not designed to serve public ends, but rather the interests of shareholders and users, in which the free flow of information is paramount and economic considerations are central.[504]

[501] Balkin (2014), p. 2298.

[502] CoE, 'Internet Governance – Council of Europe Strategy 2016–2019: Democracy, Human Rights and the Rule of Law in the Digital World' (Adopted at the 1252th Committee of Ministers' Deputies Meeting on 30 March 2016), para. 9; 'Joint Declaration on Freedom of Expression and the Internet'.

[503] MacKinnon et al. for UNESCO (2014), p. 22.

[504] Citron (2014), p. 227; Kohl and Fox (2017), p. 6.

Accordingly, they only have contractual obligations to their clients, and frequently develop their own community standards, without democratic participation. The protection of individuals against harmful speech or conduct is thus not a key concern, nor is it an obligation, as opposed to for states.

Although international human rights law does not directly regulate the acts of non-state actors, the UN Human Rights Council in 2014 adopted a resolution as a step towards a legally binding instrument on businesses and human rights.[505] Presently, human rights duties for intermediaries are delineated in soft law documents. Codes of conduct for businesses have been developed at the international level as well as in the form of collective and company self-regulation.[506] These include the UN Guiding Principles on Business and Human Rights; the Organisation for Economic Co-operation and Development (OECD) Guidelines for Multinational Enterprises and the Tripartite Declaration of Principles Concerning Multinational Enterprises and Social Policy (the MNE Declaration).[507] Nevertheless, such documents are abstract voluntary commitments, not specific to the ICT environment, the latter which may require special consideration and adaptation. Furthermore, whereas the human rights duties of states encompass measures to respect, protect and fulfil, the duties of corporations are limited to *respecting* human rights and—shared with the state—to provide remedies.[508] The obligation to respect involves a duty to act with due diligence to avoid infringing human rights, including conducting human rights impact assessments, such as evaluating the effects of software filtering systems or privacy settings.[509] This in turn requires the development and implementation of policies that take into account the potential impact of their services on human rights.[510] What is apparent through this framework is that ICTs do not have a duty to

[505] UN HRC, 'Elaboration of an International Legally Binding Instrument on Transnational Corporations and Other Business Enterprises with Respect to Human Rights' (25 June 2014) UN Doc. A/HRC/26/L.22/Rev.1.

[506] Company self-regulation entails that several companies jointly create industry codes of conduct which participants agree to abide by. It is usually in the form of terms of service. *See* MacKinnon et al. for UNESCO (2014), p. 55.

[507] UNHRC, 'Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie: UN Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework' (21 March 2011) UN Doc. A/HRC/17/31, Annex; OECD Guidelines for Multinational Enterprises, OECD Publishing (2011), <https://doi.org/10.1787/9789264115415-en> Accessed 14 March 2022; The Tripartite Declaration of Principles concerning Multinational Enterprises and Social Policy, Adopted by the Governing Body of the International Labour Office at its 204th Session (Geneva, November 1977) and amended at its 279th (November 2000), 295th (March 2006) and 329th (March 2017) Sessions.

[508] UN Guiding Principles on Business and Human Rights, Principles 11 and 26–30. The responsibility of companies to respect human rights exists independently of whether the state meets its own human rights obligations.

[509] UN Guiding Principles on Business and Human Rights, Principles 18 and 19.

[510] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 48.

protect, where obligations on interpersonal violence primarily arise, as this is exclusive to the functions of the state. However, state duties may arguably encompass state obligations to regulate the liability of intermediaries. The scope of such indirect liability is contested and will be explored further in the next section.

While calls have been made by *inter alia* the UN Special Rapporteur on the Freedom of Expression for Internet intermediaries to abide by the UN Guiding Principles on Business and Human Rights, such companies are in many respects distinct, considering the power to shape and influence the information environment, with intermediaries *de facto* acting like governments in political negotiations with states.[511] They are gatekeepers to the flow of information—as innovators, facilitators or censors—and thus fulfil the role of 'sovereigns of cyberspace'.[512] Who is permitted to participate and what is allowed to be communicated is decided by private actors. With their integral role in shaping the environment at hand—a public sphere important for democracy and the freedom of expression—they are placed in the position of political agents, arguably with a broader set of responsibilities.[513] Accordingly, the UN Special Rapporteur on the Freedom of Expression has emphasised that ICTs perform 'critical social and public functions', which must have an impact on their rights and responsibilities.[514] The codes of conduct developed by international institutions as a result have limited applicability to Internet intermediaries.[515]

Furthermore, Internet intermediaries are not only distinct from other corporations but vary greatly in 'size, sector, operational context, ownership structure or nature'.[516] Whereas the obligation to respect human rights remains regardless of such differences, the CoE has held that the means through which intermediaries meet this responsibility may vary.[517] Similarly, the UN Guiding Principles on Business and Human Rights note the need to consider context when delineating the human

[511] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 42. *See also* Broeders and Taylor (2017), p. 319; Laidlaw (2015), p. 94; MacKinnon et al. for UNESCO (2014), p. 22.

[512] Laidlaw (2015), p. 29; MacKinnon et al. for UNESCO (2014), p. 13.

[513] Taddeo and Floridi (2017), p. 3. For example, the CoE has called on Internet intermediaries to be mindful of the public service value of the services they provide and thus aim to avoid adverse effects on the rights of users. *See* CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries' (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies), Principle 2.1.4; Benedek and Kettemann (2013), p. 102.

[514] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. David Kaye' (30 March 2017), para. 47.

[515] Laidlaw (2015), pp. 112, 227.

[516] CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 2.1.2. *See* also Laidlaw (2015), p. 257.

[517] CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 2.1.2.

rights responsibilities of corporations.[518] For example, their capacity to enable or constrain different types of expression differs. This approach is also apparent in the case law of the ECtHR and in EU law, bearing in mind the type, size and business model of the corporation, in addition to viable means and the cost of controlling content for the company.[519] This is also the case in relation to gender-based violence, where the purpose and characteristics of intermediaries may affect the enabling of offences, for instance, whether involving direct perpetration, facilitation or exacerbation of violations. The degree of liability should thus vary accordingly.

Beyond international codes of conduct on businesses and human rights, states and corporations increasingly cooperate in regulating the Internet, be it voluntarily or through coercion, commonly known as co-regulation.[520] This has the advantage of being tailored particularly to the online environment. For example, the European Commission, together with major Internet intermediaries, has developed a "Code of conduct on countering illegal hate speech online", requiring the review of user notifications on alleged hate speech within 24 h and the removal or disabling of such content.[521] Meanwhile, self-regulation often occurs in the form of company terms of service, which derive from contract and commercial law, where companies have a right to require that users abide by rules in order to receive service.[522] Such self-regulation may be sanctioned by the state or developed for commercial purposes. The platforms may remove content or restrict user access if the user violates the terms of use. While domestic laws have limited reach due to jurisdictional constraints, terms of service apply globally to all users of that particular service. Consequently, in addition to being subject to domestic civil and criminal laws, Internet users are often subject to service agreements of, for example, social network companies or ISPs. Meanwhile, the terms of service often contend that the company is not responsible for comments made on the website but will exercise due diligence when notified.[523] The stated release of liability by a company was in fact taken into account by the ECtHR in *Pihl v Sweden* when assessing state obligations concerning defamation.[524]

[518] UN Guiding Principles on Business and Human Rights, Principle 14.

[519] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020); *Delfi v Estonia* (ECtHR).

[520] MacKinnon et al. for UNESCO (2014), p. 56.

[521] European Commission, 'Code of conduct on countering illegal hate speech online' (30 June 2016): <https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> Accessed 7 March 2022. *See*, also, European Commission, 'Code of Practice on Disinformation' (28 September 2018) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation> Accessed 14 March 2022.

[522] MacKinnon et al. for UNESCO (2014), p. 20.

[523] Halder and Jaishankar (2011), p. 390.

[524] *Pihl v Sweden* (ECtHR), para. 32. This is discussed further in Sect. 4.3.4.

Self-regulation by Internet intermediaries is in certain instances presumed in international sources (e.g. the EU's e-Commerce Directive), or encouraged as a possible means of protection (the EU's Child Pornography Directive).[525] Furthermore, the CoE Convention on Preventing and Combating Violence against Women and Domestic Violence (the Istanbul Convention) obliges states to '*encourage* the private sector, the information and communication technology sector and the media, with due respect for freedom of expression and their independence, to participate in the elaboration and implementation of policies and to set guidelines and self-regulatory standards to prevent violence against women and to enhance respect for their dignity'.[526] A similar provision can be found in the CoE Convention on Protection of Children against Sexual Exploitation and Sexual Abuse (the Lanzarote Convention).[527] Self-regulation is thus not an obligation for intermediaries but occasionally the *promotion* of developing terms of use is construed as a state obligation.

Nevertheless, self-regulatory regimes present certain challenges within international human rights law. The terms of service are often broadly formulated—creating an uncertainty as to what content is prohibited—inconsistently applied or enforced in a manner that amplifies existing structural inequality. This may negatively affect vulnerable groups, including women in general or specific groups of women.[528] Often the terms of service are formulated in conformity with domestic laws rather than international standards.[529] For example, the private sector has the right to set terms of service which are more restrictive of speech than what is allowed by states

[525] Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, para. 47; 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 L 178/1, para. 49. The CERD also encourages self-regulation by ISPs. *See* CERD, 'General Recommendation No. 35: Combating Racist Hate Speech' (26 September 2013), para. 42.

[526] Art. 17 (1) of the Istanbul Convention. Emphasis added.

[527] Art. 9 of the Lanzarote Convention, provides that states shall encourage the private sector, including the information and communication technology sector, to participate in the elaboration and implementation of policies to prevent sexual exploitation and abuse of children and to implement internal norms through self-regulation.

[528] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (11 May 2016) UN Doc. A/HRC/32/38, para. 52. Furthermore, the terms of use are often inconsistent in their approach to various forms of speech, for example, with social media platforms failing to restrict speech constituting hate speech, while banning speech generally considered lawful, such as adult nudity. *See* Pavan (2017), p. 67; Benedek and Kettemann (2013), p. 104. The inconsistent enforcement of policies on hate speech has been noted by the UN Special Rapporteur on the Freedom of Expression as negatively affecting minorities. *See* UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (6 April 2018) UN Doc. A/HRC/38/35, para. 27.

[529] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (11 May 2016), para. 52.

under international human rights law. Additionally, these entities generally do not have sufficiently sophisticated knowledge of either international human rights law or domestic laws to identify illegal material, which often leads to the removal of legitimate content. Especially assessments involving the appropriate balance between various rights, such as the right to privacy and the freedom of expression, are complex. The IACmHR has, for example, held that private actors 'lack the ability to weigh rights and to interpret the law in accordance with freedom of speech and other human rights standards'.[530] The delegation of censorship measures to private entities may thus lead to self-protective and over-broad restrictions.[531] Meanwhile, in practice, the terms of use of are often based on user agreements employed by large online platforms in the US, thus reflecting a generous approach to the scope of the freedom of expression.[532]

As a result, the need for aligning terms of service with international human rights law principles is increasingly raised.[533] In fact, a responsibility to adhere to the non-discrimination principle has been emphasised in multiple soft law sources. CoE recommendations on Internet intermediaries outline the responsibility to ensure that the actions of intermediaries do not have direct or indirect discriminatory effects on users, which may require that they 'make special provisions for certain users or groups of users in order to correct existing inequalities'.[534] Similarly, the UN Special Rapporteur on the Freedom of Expression has called for Internet intermediaries to relinquish the formalistic approach of treating all individuals alike and instead consider the vulnerability of certain groups to abuse and harassment.[535] This should accordingly be considered when developing or modifying Internet design, policies and products.

Furthermore, there is a general lack of communication with users on how terms of service are developed, interpreted and enforced, although companies are increasingly publishing transparency reports on the removal of harmful content.[536] According to the UN Special Rapporteur on the Freedom of Expression, the lack of transparency in the decision-making process of intermediaries '. . .often obscures

[530] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013), para. 105.

[531] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 40.

[532] De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 11.

[533] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (6 April 2018), para. 45.

[534] CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 2.1.5.

[535] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (6 April 2018), para. 48.

[536] MacKinnon et al. for UNESCO (2014), p. 12. Facebook publishes such yearly.

discriminatory practices or political pressure affecting the companies' decisions'.[537] There are thus increased demands on ICTs to be transparent with users on their definitions of unlawful material and the implications of privacy settings, as well as to give notice when information is shared with states.[538]

3.4.3.2 State Obligations and Intermediary Liability

3.4.3.2.1 Introduction

As affirmed, in international human rights law, states acquire positive obligations to create a legal framework that ensures respect for human rights, which includes the regulation of non-state actors, be it private individuals or corporations.[539] It may involve placing a duty of care on businesses. This applies regardless of the sphere in question but is heightened in certain domains and affects the scope of obligations. For example, in *Costello-Roberts v the United Kingdom*, the ECtHR held that 'the state cannot absolve itself from responsibility by delegating its obligations to private bodies or individuals', bordering on state attribution rationales.[540] This primarily concerns areas of public service, such as hospitals,[541] education [542] and psychiatric institutions,[543] that is, areas where the state has authority and assumes control. The public sphere attributes of the Internet entails that it may apply also to this setting. Where the state transfers such measures as filtering, monitoring or blocking to ISPs, be it through cooperation between states and ICTs or coercion through legislation, it may thus directly implicate the state.[544] Similarly, state interference with private

[537] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 42.

[538] According to the EU Agency for Fundamental Rights, social media 'could do more' to highlight and respond to abusive behaviour by clarifying what types of comments are considered unlawful. *See* European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 93. Similarly, the UN High Commissioner for Human Rights recommend that companies assess how their terms of service may have an adverse impact on the human rights of their users. *See* UNHRC, 'The Right to Privacy in the Digital Age: Report of the Office of the United Nations High Commissioner for Human Rights' (30 June 2014), para. 44.

[539] For example, the ECtHR in *Khurshid Mustafa and Tarzibachi v Sweden* (2011) 52 EHRR 24 held the state accountable for failing to protect individuals within its jurisdiction from the adverse effects on their rights and freedoms resulting from acts of private companies.

[540] *Costello-Roberts v the United Kingdom*, 19 EHRR 112, para. 27. This can be construed as either involving state attribution and negative obligations or positive obligations to regulate non-state actors.

[541] *Center for Legal Resources on behalf of Valentin Campeanu v Romania* App no 47848/08 (ECtHR, 17 July 2014), para. 130.

[542] *Costello-Roberts v the United Kingdom* (ECtHR); *O'Keeffe v Ireland* (2015) 59 EHRR 15.

[543] *Storck v Germany* (2005) 43 EHRR 96, para 103.

[544] Art. 5 and Art. 8 of the ILC, 'Responsibility of States for Internationally Wrongful Acts' (2001). For a discussion, *see* Land (2019).

companies has been deemed legitimate when there is compelling public interest, for example, for health and safety reasons, an argument relevant for placing obligations on ICTs to prevent harmful acts.[545] The UN Guiding Principles on Business and Human Rights also affirm that states must ensure that not only state bodies but also businesses under their jurisdiction respect human rights.[546] However, explicit treaty-based obligations for states to regulate companies are rare. Primarily the Optional Protocol to the CRC on the Sale of Children, Child Prostitution and Child Pornography[547] and the Convention on Cybercrime of the Council of Europe (the Budapest Convention)[548] affirm state obligations to provide an oversight of private companies, while the Istanbul Convention obliges states to encourage self-regulation.[549]

It should in this regard be noted that the practical ability of states to influence intermediaries depends on the nature of the latter. For example, ISPs are either state-owned, partially or fully privatised, or hybrids.[550] ISPs operate services within the jurisdiction of a state, be it equipment, resources or personnel, which entails that they must comply with domestic law, that is, are bound through laws based on territorial jurisdiction. The relationship between ISPs and states is as such characterised by mutual dependency.[551] This makes them more responsive than other intermediaries to external requests for the removal of specific content.[552] Meanwhile, web-hosting providers, search engines and social media platforms may be located anywhere and accordingly have a greater level of autonomy from the state in whose territory they operate. As a result, these companies tend to primarily adhere to their own terms of service.

In terms of state regulation of intermediary liability, it may involve either direct transgressions by intermediaries—for example, violations of data protection—or secondary liability, where they are held responsible for the conduct of users, by providing the infrastructure that enables the illegal acts.[553] Since gender-based online harm primarily is perpetrated by private individuals, the latter is the most

[545] *Fadeyeva v Russia* (ECtHR), para. 89.

[546] UN Guiding Principles on Business and Human Rights, Principle 3.

[547] Art. 3 of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography, GA Res. 54/263, Annex II, 54 UN GAOR Supp (No. 49) at 6, UN Doc. A/54/49, Vol. III (2000), entered into force January 18, 2002. This is affirmed in CRC, 'General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights' (17 April 2013) UN Doc. CRC/C/GC/16.

[548] *See* the Preamble and Art. 12 of the Budapest Convention.

[549] Art. 17 (1) of the Istanbul Convention.

[550] MacKinnon et al. for UNESCO (2014), p. 20.

[551] ibid., p. 22.

[552] Berger (2017), p. 30.

[553] Direct liability is primarily regulated through domestic law but also possible through EU law, in Art. 3(1) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, affirmed in Case C-610/15 *Stichting Brein v Ziggo BV and XS4ALL Internet BV* (2017) ECLI:EU:C:2017:456.

relevant to this topic and will be in focus in the next sections. As noted, monitoring and moderation of online content by intermediaries are mainly performed as voluntary commitments or through self- or co-regulation. Nevertheless, there is room to argue that state obligations to protect individuals or particular groups may in fact involve restricting access to online material through the regulation of Internet intermediaries. Secondary liability regimes in several respects mitigate harmful online offences. Such standards strengthen requirements of removing or restricting access to unlawful material. Blocking, filtering and censoring material may in certain instances prevent the publication or transmission of harmful content. This encompasses both communications causing individual harm, such as sexual harassment, and group-based harm, such as sexist hate speech and harmful pornography. It may also be undertaken to reduce additional harm for victims where harmful material concerning them is present online. Secondary liability rules also incentivise intermediaries to bar perpetrators and to compensate victims.

While the developing approach to secondary liability is largely cohesive, at least in the European context, the permissible scope of monitoring and moderating online content is subject to greater variation. In addition to delineating the scope of secondary liability, the following parts will thus consider two questions in relation to mechanisms of controlling content. Primarily, are such measures *compatible* with international human rights law norms, that is, do they constitute lawful interferences? Secondly, can particular measures be considered part of state *obligations* to protect individuals against harm on the Internet?

3.4.3.2.2 EU Law and Secondary Liability

Given the extensive governance of the Internet by intermediaries, legal frameworks for holding such companies accountable for content published by third parties are increasing in both EU law and international human rights law. In international human rights law, this does not involve direct liability for corporations but rather state obligations to adopt liability regimes. Various approaches to secondary liability exist, ranging from strict liability to broad immunity.[554] Nevertheless, the current approach in EU law and, tentatively, in international human rights law, is the application of safe harbour regimes. It is clear that strict liability is not viewed as an appropriate venue from a human rights law perspective, as it is deemed

[554] These include (a) strict liability (responsibility for third-party content even when the company is not aware that the content exists/is illegal); (b) safe harbour regulations (liability under certain circumstances, for example, if a company does not remove content upon notification, notify the communicator or disconnect repeat offenders), which does not require intermediaries to proactively monitor or filter content and; (c) broad immunity, that is, exemption from liability. *See* MacKinnon et al. for UNESCO (2014), p. 41.

incompatible with minimum standards of ensuring the freedom of expression.[555] Meanwhile, broad immunity would undermine the effective prevention of harmful speech.

The regulation of intermediary liability developed by the EU currently comprises sector-specific instruments on copyright infringement, terrorist content, child sexual abuse material and hate speech, covering the dissemination of such content on certain types of services (such as videosharing platforms).[556] It also involves broader frameworks, not limited to specific content or platforms. While the e-Commerce Directive governs the exemption of liability for technical intermediaries, including ISPs, search engines and social media platforms,[557] the proposed Digital Services Act (DSA) regulates intermediary responsibility.[558] These frameworks are mainly principle-based, in view of the rapid evolution of technology, and have influenced the approach to state obligations in international human rights law, for example, by the ECtHR.

The e-Commerce Directive of 2000 harmonises principles on the circumstances under which an intermediary is exempt from liability for unlawful content communicated by a third party, regardless of the nature of the liability, that is, whether civil or criminal. The Directive does thus not define what constitutes "illegal" content, consequently to be determined at the domestic level. If there were no immunity, intermediaries would be treated as publishers of content and thus held liable under, for example, domestic defamation laws. The purpose of the e-Commerce Directive was to foster the development of information society services in the internal market and, although not an inherent human rights law instrument, aimed to strike a balance between this overarching purpose with the societal interests of removing illegal information and the protection of fundamental rights.[559] The rationale for exempting liability for digital platforms is the acknowledgment of the difficulties for intermediaries in detecting and assessing the illegality of material. It is also in consideration

[555] IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (31 December 2013), para. 95.

[556] Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography; Audiovisual Media Services Directive (2018); Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA; Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

[557] 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').

[558] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020).

[559] European Commission, 'Commission Staff Working Document of 11 January 2012 on online services, including e-commerce, in the Single Market', (2012) SEC(2011) 1641, p. 24.

of the potential effects of liability regimes, such as collateral censorship.[560] This again indicates that, beyond practical issues, the value of the Internet as a vehicle for democracy is a factor limiting liability in this sphere. As noted in an EU policy report, the benefits of secondary liability regulation should not be judged solely in relation to the extent it successfully deters unlawful communication, but also the degree to which it does not dissuade lawful expressions.[561] That is, the delineation of intermediary liability is reflective of the general balancing between the freedom of expression and protection against harmful speech.

The e-Commerce Directive does not oblige states to implement a particular scheme of liability, but rather to ensure a "safe harbour" for intermediaries under specific circumstances, with the result that domestic rules on intermediary liability may vary. It establishes exemptions of liability for intermediaries where the activity is limited to the technical process of operating and providing access to a communication network over which information made available by third parties is transmitted or temporarily stored.[562] It thus draws a distinction between "active" and "passive" digital platforms.[563] The safe harbour principle is dependent on the role of the intermediary as either a conduit, cache or host, generating different exemptions of liability.[564]

ISPs generally cache material. With regard to caching, exemption is more limited but the company is not considered liable if it acts expeditiously to remove or disable access to information it has stored when having actual knowledge that information at the initial source of transmission has been removed from the network or access has been disabled, or a court or authority has ordered the removal or disablement and it was not done expeditiously.[565]

Meanwhile, hosts store information provided by a recipient of the service. This encompasses most website operators and social media networks with, for instance, Facebook categorised as a host by the ECJ.[566] Nevertheless, it will depend on the characteristics of the particular social media company and the circumstances of the case. Hosts are exempt from liability if they did not know, nor was it apparent that the information was unlawful, or if they removed it expeditiously once they

[560] EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee' (October 2017), p. 4; Balkin (2014), p. 2309.

[561] EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee', p. 12.

[562] Art. 12–14 of the Directive on Electronic Commerce (2000).

[563] Case C-324/09 *L'Oreal SA v eBay International* (2011) ECR I-6011, para. 123.

[564] The definitions of a conduit (transmits data), cache (temporarily stores information) or host (store data), are found in Art. 12, Art. 13 and Art. 14, respectively. Conduits are exempt from liability. *See* Directive on Electronic Commerce (2000), Art. 12.

[565] Art. 13 (1) (e), Directive on Electronic Commerce (2000).

[566] Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* (2019) ECLI:EU:C:2019:458, para. 22.

became aware of it.[567] In *L'Oréal SA v eBay International*, "awareness" was understood by the ECJ to mean that the intermediary, through its own investigation, or as a result of the notification of a third party, uncovers illegal activity.[568] For example, injunctions may be issued by state authorities or courts to remove or disable access to illegal information.[569] Meanwhile, social networking platforms, often with a global reach, process vast amounts of data and mainly rely on notifications by users. Other international sources provide similar restrictions on imposing liability on intermediaries. For example, the Explanatory Report to the Additional Protocol to the Budapest Convention holds that liability may be imposed on intermediaries if there is "knowledge and control" over the information which is transmitted or stored.[570]

In contrast to intermediaries, publishers—such as online news websites creating their own content—are not exempt from liability but must adhere to domestic regulations, generally placing the same form of liability on publishers as the creators of the material. Accordingly, the ECJ has concluded that the safe harbour regime does not preclude states from enacting legislation enforcing civil liability for online news outlets in cases of defamation, as these are not encompassed by the Directive.[571] This stems from the presumption that a publisher has knowledge of the information and the ability to exercise editorial control.[572] For example, the ECtHR has affirmed that the prosecution of the owner of a website publishing obscene pornography fell within the scope of the margin of appreciation of states.[573] However, in practice, such as in the case law of the ECtHR, online news portals have been

[567] Art. 14 (1) (a) of the Directive on Electronic Commerce (2000). It should be noted that states increasingly consider intermediaries to be active rather than passive and thus not exempt from liability, for example, by organising, indexing and linking material, such as on YouTube, Facebook and Google. *See* EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee', p. 24.

[568] However, notification does not *per se* remove exemption from immunity as the information may be insufficient or false. *See* Case C-324/09 *L'Oreal SA v eBay International* (2011), para. 122.

[569] Art. 14 (3) of the Directive on Electronic Commerce (2000).

[570] However, '[i]t is not sufficient, for example, that a service provider served as a conduit for, or hosted a website or newsroom containing such material, without the required intent under domestic law in the particular case'. *See* CoE, 'Explanatory Report on the Convention on Cybercrime', para. 102. *See* also CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 1.3.7.

[571] C-291/13 *Sotiris Papasavvas v O Fileleftheros Dimosia Etairia Ltd* (2014) ECLI:EU:C:2014:2209,

[572] ibid., para. 46: 'The limitations of civil liability specified in Articles 12 to 14 of Directive 2000/31 do not apply to the case of a newspaper publishing company which operates a website on which the online version of a newspaper is posted, that company being, moreover, remunerated by income generated by commercial advertisements posted on that website, since it has knowledge of the information posted and exercises control over that information, whether or not access to that website is free of charge.'

[573] *Perrin v the United Kingdom* (ECtHR).

treated differently than print publishers in relation to liability for third party content, further developed in the following section.[574]

Rather than resulting in a harmonisation of domestic practices, studies indicate that the implementation of the e-Commerce Directive has varied greatly among Member States. This is mainly due to the lack of clarity of central concepts, such as the categorisation of certain online services as "active" or "passive" and what is considered illegal content, "knowledge" and "expeditiously" removing material.[575] The Directive has as a result been criticised for incentivising the immediate removal of content by intermediaries upon notification.[576] As the economic goals of intermediaries are generally unaffected by the removal of information, while sanctions for enabling unlawful communication may be significant, there is an inclination to remove speech even when not clearly unlawful.[577] In view of such gaps and given the rapid development of new technologies and fundamental social changes since the e-Commerce Directive, the proposed DSA aims to provide a more up-to-date approach.

The draft DSA, in contrast to the e-Commerce Directive, establishes the concept of intermediary responsibility and sets up a due diligence regime. This includes responsibilities for platforms to publish annual content moderation reports, to create an internal complaint-handling system, and to suspend users that frequently post illegal content.[578] A distinction is in this regard made between "illegal" and "harmful" online content.[579] Such a delineation is increasingly employed also at the domestic level and in doctrine.[580] Whereas the Directive proposes stringent rules on the regulation of illegal content, such as expeditiously removing material through notice-and-takedown mechanisms, responsibilities *vis-à-vis* harmful content solely extends to "risk mitigation measures" for Very Large Online Platforms (VLOPs).[581] While such mitigation efforts are to be developed through self- and co-regulatory agreements, the particular means of addressing such content is not specified.

[574] *Delfi v Estonia* (ECtHR).

[575] European Parliamentary Research Service, 'Reform of the EU Liability Regime for Online Intermediaries: Background on the Forthcoming Digital Services Act' (May 2020), pp. 5–6

[576] EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee', p. 10. *See* also Laidlaw (2015), p. 127.

[577] EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee', p. 13.

[578] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020).

[579] ibid., Art. 2 (g). Cf. ibid., para. 5 and para. 68.

[580] *See,* for example, the Canadian government: <https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html> Accessed 15 March 2022; Bonnici and de vey Mestdagh (2005).

[581] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020), para. 68.

Suggestions of tools to prevent harmful material includes improving media literacy skills and making available parental control tools and rating systems.[582]

In the DSA, "illegal content" is defined as '. . .any information, which, in itself or by its reference to an activity, including the sale of products or provision of services is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law'.[583] As there is no broad harmonisation within the EU on the criminalisation of online content, "illegal" content currently concerns racist and xenophobic hate speech,[584] child sexual abuse images,[585] infringements of intellectual property rights,[586] and terrorist activities.[587] This is construed as content 'with the highest potential negative impact on the society', that is, the individual and social harm is deemed particularly severe.[588] This is aligned with international treaties comprising obligations to prohibit such forms of speech.[589] Meanwhile, the unlawful non-consensual sharing of private images and online stalking are mentioned as examples of illegal acts regulated in certain Member States.[590]

In contrast, "harmful" online content is not defined in the proposal. Nevertheless, it is generally construed as content or conduct that does not individually rise to the level of a criminal offence, or where the legality varies across states but may nevertheless be harmful, for example, from the perspective of gender equality.[591] Such harm is viewed as being relative to the particular characteristics of the individual or group, such as sex, ethnicity, age and culture. It is thus perceived as a primarily subjective harm.[592] This is currently considered to include harassment,

[582] De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 78.

[583] Art. 2 (g), Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020).

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

[585] Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography.

[586] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020), para. 12.

[587] Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism.

[588] De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 12.

[589] For example, Art. 4 of the ICERD, Art. 34 of the CRC; OP to the CRC on the sale of children, child prostitution and child pornography; Council of Europe Convention on the Prevention of Terrorism, CETS No. 196, 16 May 2005.

[590] Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020), para. 12.

[591] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, 'Illegal and Harmful Content on the Internet' (1996) 1996COM(96) 487, p. 11.

[592] ibid., p. 11.

bullying, promotion of self-harm or eating disorders, and mis- or disinformation.[593] Although the EU has taken measures to prevent gender-based violence, there is currently limited harmonisation in this area.[594]

Nevertheless, the EU Gender Equality Strategy 2020–2025 includes an aim to extend the scope of harmonised crimes within the meaning of Article 83(1) of the Treaty on the Functioning of the European Union (TFEU) to certain forms of gender-based violence.[595] In view of this, the European Commission in March of 2022 proposed a directive harmonising the criminalisation of violence against women that falls within the remit of EU law, including rape and certain forms of cyber violence, namely image-based sexual abuse, harassment and incitement to violence or hatred.[596] The objective of the proposal is to align EU law with international standards and more effectively combat violence against women, as gender equality is a core value of the EU.[597] The Istanbul Convention is in this regard the most prominent source of reference.[598] While this is a significant step towards regulating gender-based online violence, in view of the narrow definitions of certain offences,[599] there is still a risk that various forms of online gender-based harm will mainly be categorised as "harmful", meaning that the means of prevention are limited and mainly placed on the user, rather than on intermediaries.

Beyond the perceived gravity of offences, the distinction between "illegal" and "harmful" material is also in part practical, considering means of intermediary control. The two categories may call for different types of legal and technological responses.[600] If content is "harmful", it may be sufficient to restrict access for vulnerable groups, such as children, to place a duty of care on intermediaries to label the content as harmful or to encourage users to apply filters. From a practical

[593] European Parliamentary Research Service, 'Reform of the EU Liability Regime for Online Intermediaries: Background on the Forthcoming Digital Services Act' (May 2020), p. 10.

[594] European Parliament resolution of 26 October 2017 on combating sexual harassment and abuse in the EU (2017/2897(RSP)); Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, 'A Union of Equality: Gender Equality Strategy 2020–2025', COM/2020/152.

[595] Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, 'A Union of Equality: Gender Equality Strategy 2020–2025', COM/2020/152; Consolidated Version of the Treaty on the Functioning of the European Union (TFEU) (2016) OJ C202/1.

[596] Art. 5, Art. 7, Art. 9 and Art. 10 (*see* also Art. 8 on cyber stalking) of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence' (8 March 2022).

[597] Mainly, Art. 2 of the Treaty on European Union and Art. 21 and Art. 23 of the Charter of Fundamental Rights of the European Union.

[598] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence' (8 March 2022), p. 3.

[599] For example, cyber harassment and rape. *See* further discussion in Sects. 4.2 and 4.3.

[600] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, 'Illegal and Harmful Content on the Internet' (1996) 1996COM(96) 487, p. 10.

standpoint, the level of precision in detecting illegal material in current technology may also vary depending on the content. Studies indicate that algorithms are relatively effective in identifying child abuse images.[601] In contrast, where content requires contextual assessments and consideration of nuance, such as hate speech, they are not as effective but may rather necessitate human moderation. This is even more complex in the case of "harmful" content, where intermediaries often have to perform a complex balancing between the freedom of expression and the rights affected by the harmful content. As a result, large online platforms have called for a focus on the *ability* to control illegal content.[602] There is thus a risk that technology in this regard drives the determination of what is considered "illegal", that is, that practical aspects in controlling certain forms of speech affect the hierarchy of offences and intermediary obligations.

It should be noted that there is no such delineation between different forms of harmful acts in international human rights law. Rights involve the protection against harmful acts and materials, while the severity of harm may be reflected in the right applied and the content of obligations. In contrast to EU law, states incur a range of obligations—from providing education on gender stereotypes to criminalisation—which diminishes the need for such a distinction. Thus, whereas pornography promoting gender stereotypes would most likely be categorised as "harmful" online content in view of the DSA, obligations to eliminate such arise for state parties under various human rights law treaties, for example, in Article 5 of the United Nations Convention of the Elimination of All Forms of Discrimination (CEDAW). However, such obligations are not as far-reaching as for content categorised, for example, as sexual violence, and a hierarchy of prohibited conduct is implied in practice. Nevertheless, international human rights law may influence harmonisation efforts in the EU and beyond, extending the scope of "illegal" online content and providing definitions of the same. At the same time, as will be discussed in the next section, the ECtHR has made a distinction between "clearly unlawful" content and other forms of speech that may be restricted in the context of the Internet.

Furthermore, the proposed DSA takes a differentiated approach to intermediary liability, considering that the level of harm and the identity of the harmed party must be weighed against the cost of prevention, in view of the size of the intermediary and their economic and social impact.[603] For example, VLOPs will be required to assess

[601] CoE, 'Respecting human rights and the rule of law when using automated technology to detect online child sexual exploitation and abuse: Independent experts' report' (June 2021) <https://rm.coe.int/respecting-human-rights-and-the-rule-of-law-when-using-automated-techn/1680a2f5ee> Accessed 15 March 2022; European Parliament, European Parliamentary Research Service, 'Curbing the surge in online child abuse: the dual role of digital technology in fighting and facilitating its proliferation' (2020) <https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/659360/EPRS_BRI(2020)659360_EN.pdf> Accessed 15 March 2022.

[602] EDiMA, 'Fundamentals of the Online Responsibility Framework Series: A Legal Basis to Act' <https://www.euractiv.com/wp-content/uploads/sites/2/2020/10/ORF-Series_-Basis-to-Act_EDiMA.pdf> Accessed 15 March 2022.

[603] De Streel et al. (2018), p. 42.

systemic risks of the operation and use of their services on a yearly basis, including misuse and the impact of their services on fundamental rights.[604] This may require them to adapt the design of, for example, content moderation and algorithmic detection systems.[605] Meanwhile, costs of prevention arise, for example, through human monitoring and the development of detection software, which may be substantial for smaller corporations. VLOPs are deemed to have greater financial and technological resources to remove illegal material, whereas liability rules may prevent small platforms from entering the market by requiring cumbersome and costly measures. As noted above, considering secondary liability relative to the size, nature, function and organisational structure of intermediaries is emerging as the preferred approach in multiple sources, including the Audiovisual Media Service Directive and by the CoE.[606] It is also in line with fair balancing exercises by human rights courts in conflicts of rights, evident in the case law of the ECtHR.[607]

3.4.3.2.3 The European Court of Human Rights and Secondary Liability

3.4.3.2.3.1 Media Publishers

The ECtHR has in a limited number of cases reviewed online media corporations and secondary liability. *Delfi AS v Estonia* is a central case in this regard.[608] The applicant company argued that its freedom of expression had been violated in that it had been held liable for third-party comments posted on its Internet news portal. Delfi was one of the major news portals in Estonia and published up to 330 news articles a day. Commentary sections were provided in conjunction with news articles. The comments were uploaded automatically without editing or moderation by the company. The articles received approximately 10,000 comments daily, mainly posted under pseudonyms. The website had a notice-and-takedown system where readers could mark comments as insulting or as inciting hatred, upon which the comments would be removed expeditiously. There was also automatic deletion of comments including certain obscene words. Delfi posted "Rules of Comment" on its website, detailing these mechanisms.

---

[604] Art. 26, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020).

[605] ibid., Art. 27.

[606] The content of "appropriate measures" is to be determined in light of '. . .the nature of the content in question, the harm it may cause, the characteristics of the category of persons to be protected as well as the rights and legitimate interests at stake, including those of the video-sharing platform providers and the users having created or uploaded the content as well as the general public interest'. *See* Art. 28b (3) of the Audiovisual Media Services Directive (2018); CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principles 1.1.5 and 1.3.9.

[607] For an overview, *see* also Smet (2010); Angelopoulos and Smet (2016).

[608] *Delfi v Estonia* (ECtHR).

Delfi published an article in January 2006, attracting 185 comments, of which approximately 20 involved personal threats and offensive language against an individual implicated in the article. These included anti-semitic slurs, calls for sending him into an oven and that he should go drown himself or be lynched. In March of the same year, the victim's lawyers requested that Delfi remove the offensive comments. Delfi deleted the statements on the same day as the request, approximately 6 weeks after their publication. The victim nevertheless requested compensation for non-pecuniary damage, which was awarded by the domestic court.

As a first step, the Court reviewed CoE declarations and EU directives on this issue and concluded that such instruments distinguish between online media and traditional publishers in terms of liability. A similar approach was adopted by the Court, which argued that because of the 'particular nature of the Internet', the duties and responsibilities placed on Internet news portals differ from those of traditional publishers in terms of third-party content.[609] Thereafter, the ECtHR assessed the nature of the remarks. Both the domestic court and the ECtHR categorised the comments as incitement to hatred or violence, which were "clearly unlawful", violating the right to privacy of the person.[610] It concluded that '. . .the establishment of their unlawful nature did not require any linguistic or legal analysis since the remarks were on their face manifestly unlawful'. [611] It thus examined the issue of liability from this standpoint. Firstly, the aim by the domestic court in restricting the freedom of expression of the company was legitimate, as it concerned the protection of the rights and freedom of others. The Court noted that the Internet can be both harmful and beneficial, but '. . .the risk of harm posed by content and communications on the Internet to the exercise and enjoyment of human rights and freedoms, particularly the right to respect for private life, is certainly higher than that posed by the press'.[612] It held that by virtue of Article 17, certain speech incompatible with the values of the ECHR is not protected under Article 10. As noted previously, this has in practice been limited to hate speech and incitement to violence against, primarily, ethnic or religious groups. As the comments constituted hate speech, the authors were thus not protected by the freedom of expression. The central question was accordingly whether the liability for third-party comments placed on the company violated its right to disseminate information.

In order to assess the proportionality of the interference, the Court reviewed (1) the context of the comments; (2) the liability of the authors of the comments; (3) measures taken by the applicant company and (4) consequences for the company.[613] As regards point (1), the news portal was one of the biggest media publications in the country with a wide readership and general awareness of the controversial nature of the comments it tended to attract. The Court also assessed the

[609] ibid., para. 113.
[610] ibid., para. 115.
[611] ibid., para. 117.
[612] ibid., para. 133.
[613] ibid., paras. 144–161.

ability of the company to control content on its website. Although the article itself was neutral, the website was designed in such a way that it invited visitors to add their own opinions to news articles and the company had an economic interest in the posting of such comments. The company was considered to have exercised a substantial degree of control over the comments published on its portal. The Court also considered how the case related to the e-Commerce Directive. Although Delfi argued that it was an intermediary, protected by the "safe harbour" provisions, the domestic court, with which the ECtHR agreed, held that it did not act as an intermediary but as a media publisher, due to its degree of editorial control, and could thus not benefit from such standards.[614] The relevant factor was whether the website operator was 'merely technical, automatic and passive in nature', referring to the e-Commerce Directive.[615]

In relation to point (2), the Court considered user anonymity on the Internet, which serves the purpose of ensuring the free flow of ideas and information. Additionally, it is often not feasible to hold the authors of the comments accountable, as practice in Estonia had demonstrated. As for point (3), the Court considered whether it was reasonable to place liability on the company as a result of a failure to remove a comment on its own initiative during a period of 6 weeks, even though it was deleted once notified. In this case, the company had a disclaimer, an automatic system of deletion of comments containing certain words and a notice-and-takedown system. It did thus not neglect its duty to reduce the harm suffered by victims. In this regard, the Court expounded generally on the harm of online comments. It discussed the fact that, in relation to certain offences, there may be no individual victims, for example, in situations of insults against a particular group. Furthermore, in cases of individual harm resulting from hate speech, a person has a more limited ability to continuously monitor the Internet than large commercial Internet news portals, in order to remove such comments. The Court held that a notice-and-takedown approach was consequently insufficient. Accordingly,

> . . .in cases such as the present one, where third-party user comments are in the form of hate speech and direct threats to the physical integrity of individuals, as understood in the Court's case-law. . .the Court considers. . .that the rights and interests of others and of society as a whole may entitle Contracting States to impose liability on Internet news portals, without contravening Article 10 of the Convention, if they fail to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties.[616]

The same form of liability would not extend to speech requiring a more nuanced legal assessment and harm of a lesser gravity. Although indicating in the *obiter dictum* that also defamation is "clearly unlawful", a different approach has been taken in practice.[617] The dissenting judges also criticised the fact that the majority

[614] ibid., paras. 112–113.

[615] ibid., para. 128.

[616] ibid., para. 159.

[617] ibid., para. 110.

failed to categorise the comments and analyse the extent to which they constituted real threats.[618]

The Court did not imply liability for website operators or social media networks that do not create any content, but only those that actively encourage and moderate comments.[619] This indicates a broader standard of freedom of expression on unmoderated platforms, such as social media, personal blogs and websites run as hobbies. Nevertheless, the possibility of holding the authors of comments responsible remains in those instances.

Furthermore, the Court did not explicitly require pre-monitoring of comments which, according to concurring opinions, could lead to a disproportionate interference, but rather to remove unlawful content without delay.[620] However, the effect of the judgment is equivalent to prescribing pre-monitoring as the Court held that a notice-and-takedown system was not sufficiently effective. According to dissenting Judge Sajó and Judge Tsotsoria, the ruling may lead to self-censorship, prior restraint and incentives to discontinue the comments feature, as companies generally do not have the resources or legal knowledge to pre-screen all comments.[621] Other consequences may be real-name registration policies undermining anonymity. While this diminishes the democratic force of the Internet by restricting the freedom of expression, more extensive measures to prevent harmful speech may arguably enhance the diversity of opinions.

The case of *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (*MTE and Index v Hungary*) concerned two companies which operated websites that allowed users to comment on their articles.[622] Comments could be uploaded following registration and were not previewed, edited or moderated by the applicants. The MTE website published an opinion piece criticising two real estate management websites, alleging unethical practices. This attracted comments criticising the companies mentioned. The other website—Index—republished the opinion, which also drew irate comments. Upon notification of this, the applicants removed the comments immediately. Nevertheless, the company operating the real estate management websites brought a civil suit in the domestic court, arguing that the opinion piece and the comments had tarnished its good reputation. Index argued that, as an intermediary, it was not responsible for user comments. However, the domestic court did not consider the companies to be intermediaries and MTE and

[618] ibid., Joint dissenting opinion of Judges Sajó and Tsotoria, paras. 12–14.

[619] ibid., para. 116.

[620] ibid., Concurring opinion of Judges Raimondi, Karakas, De Gaetano and Kjolbro, para. 6.

[621] ibid., Joint dissenting opinion of Judges Sajó and Tsotsoria, para. 8. *See* also critique in UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (22 May 2015), para. 54.

[622] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR).

Index were both found responsible on the basis of objective liability for the user comments, by having enabled the function of open commentary.[623]

The Court reiterated the principles established in *Delfi v Estonia* for assessing the liability of media publishers. First, the Court noted the different characteristics of the two companies at hand. One was a major news portal, run on a commercial basis and which attracted a large number of comments (Index). The other was an association of Internet content providers, which did not attract user comments to the same degree (MTE). However, both companies were approached in a similar manner, apart from consideration of the particularly severe economic consequences of sanctions for the latter company.[624] Regarding the content, the Court held that '[a]lthough offensive and vulgar, the incriminated comments did not constitute clearly unlawful speech; and they certainly did not amount to hate speech or incitement to violence'.[625] The Court did not consider the comments defamatory, but rather value judgments and opinions, albeit offensive and generating reputational harm. It reiterated that vulgarity is not unlawful, as the style of communication is protected alongside content, noting that the commonly offensive nature of speech on the Internet reduces its impact and thus the level of harm.[626]

The Court found that the liability standard of Hungary failed to balance the competing rights of the parties involved. The applicants had taken measures of protection, which included disclaimers and notice-and-takedown systems. Meanwhile, the injured parties had not requested that the companies remove the information. Holding the companies liable on the basis that they should have foreseen remarks in breach of the law by allowing unfiltered comments sections, according to the Court required '. . .excessive and impracticable forethought capable of undermining freedom of the right to impart information on the Internet'.[627] The consequences of maintaining an objective liability could lead to a closing of the comments section and thus cause a chilling effect on the freedom of expression on the Internet. The main difference between this case and *Delfi v Estonia* was hence the nature of the comments.

In the recent case of *Sanchez v France*, the ECtHR accepted criminal sanctions against a politician for not expeditiously removing third-party comments to a post he made on Facebook.[628] The post criticised his political opponent in the run-up to the French Parliamentary elections, which attracted comments amounting to hate speech against Muslims. These comments were present online for approximately 6 weeks.

[623] Although the ECtHR did not explicitly expound upon whether the companies were intermediaries in terms of EU law, it did not challenge the domestic court's ruling that the companies did not qualify for liability exemption and were to be approached as publishers, in consideration of the control they had over the comments section.

[624] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 88.

[625] ibid., para. 64.

[626] ibid., para. 77.

[627] ibid., para. 82.

[628] *Sanchez v France* App no 45581/15 (ECtHR, 2 September 2021).

The Court, in balancing the opposing rights, applied the criteria set out in *Delfi v Estonia*, thus extending the scope beyond online commercial media publishers. The applicant was in this regard also considered a publisher by having set up a public communication service and willingly making the account public, allowing friends to post comments. This was done in connection to an election campaign and thus aimed to reach a broader public. Additionally, in view of the nature of the posts, the Court considered that he could not have been unaware that they may attract hostile comments. The particular duties for politicians in the exercise of their freedom of expression was noted, including combating hate speech. In this regard, the Court emphasised the shared responsibility between online platforms, publishers and users *vis-à-vis* unlawful comments, requiring the monitoring of content. Nevertheless, the Court did not discuss whether this liability ensues upon knowledge of the comments or, in effect, requires pre-monitoring, in line with *Delfi v Estonia*. Arguably, whereas commercial platforms may have the resources to continuously monitor websites, individual users may not. Nevertheless, given that the Court has indicated that bloggers and influential social media users may similarly fulfil the role of a "public watchdog", a more expansive approach to liability for users appears to be developing.[629]

3.4.3.2.3.2 Intermediaries

The ECtHR case of *Payam Tamiz v the United Kingdom* concerned defamation published on a blog.[630] Google Inc. was a corporation registered and with its principal place of business in the United States. Google had a blog-publishing service—Blogger.com—which allowed Internet users to create blogs free of charge. Its "Content Policy" set out restrictions for the content on blogs, including prohibiting child abuse images and the promotion of racial hatred. Additionally, Google Inc. included a "Report Abuse" feature, which enabled the reporting of defamation, libel and slander. However, it made clear that it operated under US laws on defamation and would only remove such material if found libelous by a domestic court. This was due to the vast volume of content uploaded daily, making it impractical for the corporation to continuously review material.

Several defamatory comments against Mr. Tamiz were added to a blog article. The applicant used the "Report Abuse" function but did not receive a reply. Several months later he sent a letter of claim to Google Inc., which in turn replied that it would not remove the content but forwarded the letter to the publisher of the blog who removed the webpage. The applicant subsequently brought a libel claim against Google Inc. in a UK court, which did not find in favour of the applicant.

The ECtHR held that the domestic court had sufficiently considered the relevant elements of the nature and context of the comments, the action taken by Google Inc.

---

[629] For example, *Magyar Helsinki Bizottság v Hungary* App no 18030/11 (ECtHR, 8 November 2016), para. 168.

[630] *Payam Tamiz v the United Kingdom* (ECtHR).

subsequent to notification, possibilities of holding the author of the comment liable and the consequences for the individual *versus* the provider of the platform. [631] The Court agreed that the applicant was not without recourse. For example, he could have brought libel proceedings against the authors of the comments. Although the ECtHR noted the practical difficulties in doing so due to the anonymity of authors, it argued that an application to seek disclosure of the identity was possible. However, even the domestic court judge argued that this possibility '. . .may be regarded as more theoretical than real. . .'. [632] The applicant could also have brought proceedings against the publisher of the blog. The Court differentiated the case from *Delfi v Estonia* in that Delfi was a major, professionally managed Internet news portal run on a commercial basis, which invited user comments to its own news articles. The same liability could not be imposed on social media platforms where the platform provider did not offer its own content, that is, implying liability exemption extending to hosting services.[633]

In having reviewed several international sources of relevance, the Court found that the domestic ruling was in line with the general position in international law accepting limited intermediary liability for third-party content, if they fail to act expeditiously in removing or disabling access once it has become aware of the illegality of the content.[634] In this case, the Court considered that the content was removed within an acceptable time-frame.[635] The Court emphasised the important role of intermediaries such as Google in '. . .facilitating access to information and debate on a wide range of political, social and cultural topics', generating a wide margin of appreciation for the state, finding that a fair balance had been struck.[636] In the balancing of interests, the Court thus in effect noted the important democratic function of such intermediaries, as well as the Internet in general, in addition to its distinctive technological features.

In the case of *Pihl v Sweden*, the applicant complained that a blog post had been published accusing him of being involved in a Nazi party, with an additional comment claiming that he was a hash-junkie.[637] The blog was small and run by a non-profit association. The blog allowed comments to be posted and it stated clearly that such comments were not checked before publication and that commentators were responsible for their own statements. Commentators were requested to 'display good manners and obey the law'. [638] Upon noticing the blog post 9 days later, the applicant wrote a comment stating that the information was incorrect and requested

[631] ibid., para. 87.

[632] ibid., para. 28.

[633] ibid., para. 85.

[634] ibid., paras. 54–56, 84.

[635] The removal occurred approximately 4 months after the applicant reported the comments, albeit Google Inc. did not agree with the applicant on the date on which it had been informed.

[636] *Payam Tamiz v the United Kingdom* (ECtHR), para. 90.

[637] *Pihl v Sweden* (ECtHR).

[638] ibid., para. 3.

that the post be taken down. The following day the blog post and comment were removed and a new blog post by the association appeared, stating that the previous post had been inaccurate. However, the applicant argued that the old post and comments were still accessible on the Internet. The applicant sued the association for defamation on the basis that it had failed to remove the post immediately. While the domestic court agreed that the comment was defamatory, the domestic law did not include defamation as a ground on which to hold an electronic platform responsible. Rather, this was limited to content such as agitation against a national ethnic group (equivalent to hate speech) and child abuse images. The domestic court did thus not find in favour of the applicant.

At the time of the review by the ECtHR, the blog post itself was subject to domestic proceedings and thus the Court solely considered the comment attached to the blog post. The ECtHR accepted the categorisation of the comment as defamatory by the domestic court.[639] The fact that the comment was offensive but did not constitute hate speech or incitement to violence was of importance.[640] The Court noted that the comment was unconnected to the post, in that it did not concern his political views. Thus, it could not have been anticipated by the association. Additionally, it concerned a small non-profit association, unknown to the wider public, making it unlikely that '. . .it would attract a large number of comments or that the comments about the applicant would be widely read'.[641] The Court noted that the association clearly stated that it did not check comments, it requested commentators to display good manners, and the post was removed a day after notification. The applicant had also taken few steps to identify the author.

In view of the fact that the blog post and comment were still accessible through search engines, the Court referred to the possibility of requesting that search engines remove information, that is, the right to be forgotten, discussed in the following section. It again noted that liability for third-party comments may have negative consequences for comment-related environments on Internet portals and thus a chilling effect on the freedom of expression online.[642] Consequently, the domestic court had struck a fair balance. It should be noted that the Court did not require an effective notice-and-takedown system. Since the post was removed quickly, this was not an issue in the case. The Court thus left the question open whether other measures than removal might be appropriate, such as rectification or a right of reply.[643]

Høiness v Norway involved a claim of negligence by the operator of an electronic discussion board to expeditiously remove speech perceived by the applicant as sexual harassment.[644] The applicant was a well-known lawyer and actively involved

[639] ibid., para. 25.

[640] ibid., para. 37.

[641] ibid., para. 31.

[642] ibid., para. 35.

[643] Discussed by Voorhoof (2017).

[644] *Høiness v Norway* App no 43624/14 (ECtHR, 19 March 2019).

in public debate. The discussion forum in question only contained user-generated content, that is, no editorial material, although accessed through the website of an online newspaper. There was no requirement to register in order to post comments and users could remain anonymous. A forum thread regarding the applicant was created, which included anonymous comments on the applicant's appearance, gossip on her sex life and sexual innuendo.[645] In terms of complaint mechanisms, there was a link next to each post where users could click to indicate that it was inappropriate. Moderators also monitored content, although it was not considered a particularly effective mechanism. Several of the comments in the case had not been noticed by moderators. However, these were removed once the operator was notified.

The Court applied its previous approach in balancing the right to privacy and the freedom of expression. While the applicant argued that the comments involved sexual harassment and a form of sex discrimination—excluded from protection under Article 10—the ECtHR considered that it was 'not necessary to examine in depth the nature of the impugned comments, as they in any event did not amount to hate speech or incitement to violence. . .'. [646] Again, it accepted that the applicant would 'have faced considerable obstacles' in pursuing claims against the anonymous commentators.[647] Nevertheless, in consideration of these facts, the ECtHR held that the state had acted within its margin of appreciation in balancing the rights and exempting the company from liability. This again affirms the distinctive approach *vis-à-vis* hate speech.

The Court in the most recent case on this issue—*Jezior v Poland*—considered that holding a local politician liable for third-party comments, amounting to defamation, violated his freedom of expression.[648] The applicant administrated a blog, primarily of interest to citizens of the town in which he resided, which allowed third-party comments without registration. The applicant occasionally monitored user comments and a notification system was in place. Rules of appropriate conduct were posted on the blog. A comment defaming the current mayor, implying his involvement in criminal activities, was posted during the pre-election period. This was not a reaction to the applicant's own posts or in respect of his views. The comment, and re-postings, were immediately removed. The applicant subsequently introduced a mandatory registration system. Nevertheless, the applicant was ordered by the domestic court to pay damages to the victim. The ECtHR concurred that the statements were defamatory but considered the measures taken by the applicant to be sufficiently effective. Requiring pre-monitoring would be excessive and undermine the freedom of expression online. In applying the same criteria as in *Delfi v Estonia*, a distinction was thus made in that it did not involve hate speech.

[645] For example, "If I were to s—her, it would have to be blindfolded. The woman is dirt-ugly – looks like a wh—".

[646] *Høiness v Norway* (ECtHR), para. 69.

[647] ibid., para. 70.

[648] *Jezior v Poland* App no 31955/11 (ECtHR, 4 June 2020).

3.4.3.3 Monitoring and Moderating Content

The liability regimes developed by the EU and the approach of the ECtHR indicate, although sparingly, certain standards in terms of monitoring and moderating content, that is, whether particular measures are compatible with or considered obligations in these areas of law. New types of monitoring and controlling content are possible on the Internet. Pre-Internet forms of regulating the freedom of expression centred primarily on fines, injunctions and the use of *post hoc* penalties.[649] Prior restraint was limited, mainly performed through injunctions, regulation of access to public spaces or control of publishers and broadcasters. Meanwhile, *ex-ante* prevention of illegal communication (e.g. content filtering or blocking) and *ex-post* measures (e.g. content removal or account termination) are possible on the Internet.[650] Regulation and control of speech are also increasingly considered when *designing* digital infrastructure.[651] Freedom of expression censorship may, for example, be embedded in network equipment, local software, consumer electronics and mobile telephone devices. Prevention capacities are consequently more extensive in cyberspace.

The monitoring of content contravening domestic law or terms of service includes the manual screening by human moderators and users of user-generated content submitted or posted on online platforms. It may occur prior to the posting of content or when content is flagged. Monitoring can also be performed through the use of automated filters that constrain publication or flag certain content on the basis of keywords or images.[652] Filtering systems may be employed by the state, ISPs and users. Increasingly, intermediaries employ software to identify proscribed material, such as hate speech. As a result of varying domestic laws, some Internet companies also use country specific filters on the basis of geolocations. Users themselves can likewise purchase and activate filtering software aligned with their moral views, thus reducing state authority in controlling information. Such self-regulatory measures are considered especially appropriate in relation to harmful content or material that is solely unsuitable for children, such as pornography.[653] This also includes parental control systems and age verification software. Additionally, users or pre-set filtering systems can rate, label and categorise online content. Nevertheless, both labelling and age verification software are generally considered ineffective measures. The former is mainly performed *ex post* publication and is challenging in view of the

[649] Regulation was mainly directed at (a) speakers; (b) spaces; and (c) predigital technologies of mass distribution. Thus, the state could arrest and prosecute individuals, control access to public spaces for assembly and monopolise, destroy or regulate means of publication and transmission, such as printing presses, books and broadcasting organisations. *See* Balkin (2014), p. 2306.

[650] ibid., p. 2318.

[651] Puddephatt for UNESCO, 'Freedom of Expression and the Internet' (2016), p. 20.

[652] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 29.

[653] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, 'Illegal and Harmful Content on the Internet' (1996) 1996COM(96) 487, p. 10. *See,* also, Akdeniz (1997).

sheer volume of online content and variations in cultural sensitivities.[654] Meanwhile, age verification systems are easily circumvented and privacy concerns arise through the collection of information.[655]

In terms of the compatibility with human rights, certain forms of monitoring involve a higher risk of collateral censorship, that is, excessive restrictions of the freedom of expression, as well as invasions on the right to privacy, for example, through surveillance and collection of personal information. In the EU, monitoring is subject to the General Data Protection Regulation (GDPR) and the complementary e-Privacy Directive, regulating the collection and processing of personal data.[656] Meanwhile, the e-Commerce Directive, the DSA and the Additional Protocol to the Budapest Convention hold that states cannot impose obligations on Internet intermediaries involving general surveillance or to actively search for illegal content.[657] A similar approach has been adopted by various UN bodies.[658] Nevertheless, this solely concerns monitoring measures imposed by authorities, and not voluntary measures by intermediaries. A requirement for ISPs to install filtering software in order to conduct blanket searches for unlawful content would thus amount to an infringement of data protection rules, since it may block also lawful communication. However, states may impose specific monitoring obligations. In *Eva Glawischnig-Piesczek v Facebook Ireland Limited*, the ECJ held that in a situation of a genuine

[654] Lievens (2010), p. 246.

[655] ibid., p. 250.

[656] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data; Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) OJ L 201/31. At the same time, recital 40 of the e-Commerce Directive provides that immunity 'should not preclude the development and effective operation. . .of technical systems of protection and identification and of technical surveillance instruments made possible by digital technology', leaving the scope of the restriction on general filtering unclear.

[657] The e-Commerce Directive prohibits member states from imposing general obligations of surveillance on ISPs over the information they transmit or store. Member states are also prohibited from imposing on them a duty to actively search out illegal content which they might host or transmit. *See* also Art. 15 of the Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (ETS No. 189) 28 January 2003; CoE, 'Explanatory Report on the Convention on Cybercrime', para. 102; Art. 7 of the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020). State authorised filtering was addressed in Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* (2012) ECLI:EU:C:2012:85 of the CJEU, concerning obligations placed on hosting provider Netlog, balancing Art. 15 and Art. 16 of the e-Commerce Directive with Art. 8 and Art. 11 of the Charter of Fundamental Rights of the EU. Additionally, in Case C-70/10 *Scarlet Extended SA v SABAM* (2011) ECR I-11959, Scarlet being a conduit provider, the CJEU held that an obligatory filtering system would involve a very high cost, borne by the intermediary.

[658] *See*, for example, UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 43.

risk of subsequent reproduction of defamatory content, an obligation to monitor and remove such specific information could be placed on Facebook without contravening the prohibition on imposing general obligations to monitor content, as described in the injunction.[659] Otherwise victims would have to institute multiple proceedings. Furthermore, as viewed in *Delfi v Estonia*, more extensive obligations on pre-publication monitoring may be placed on media publishers in relation to hate speech. It should, however, be noted that searches for specific content frequently involve general monitoring in order to detect unlawful material.

In view of such limitations, notice-and-takedown systems are increasingly the preferred methods of content moderation and the suggested forms of intermediary control by the CJEU, the DSA and the ECtHR.[660] Such processes are also commonly included in self-regulatory agreements. As active monitoring by intermediaries is onerous and subject to regulation, speech-based offences are under such systems not removed unless the company receives notice thereof.[661] Automated tools, trusted flaggers and user notifications may be employed as components of a notice-and-takedown mechanism. However, the particular means of monitoring involved in notice-and takedown regimes have different effects on user rights. While user notifications avoid general monitoring, there is a risk of both under- and over-removal of lawful speech, reflective of broader social structures. As noted previously, gender-based harm is frequently overlooked. Notifications by users may also work better in instances of individual harm, as opposed to social harm, where individuals may not feel compelled to notify the platform. This may require a more active role for intermediaries.[662]

Meanwhile, in a study of large digital platforms in the EU, automated tools were considered the most effective means of detecting illegal material.[663] However, the accuracy of automated tools depends on the type of content.[664] For example, effective automated filters have been developed to identify and remove child abuse

[659] Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* (2019), para. 41. It constituted monitoring 'in a specific case'. The CJEU was requested to consider whether it was compatible with Art. 15 of the Directive to place obligations on host providers to ensure the removal of republications of unlawful material, such as defamatory comments. According to the ECJ, Facebook was a host provider that had knowledge of illicit material on its website and did not act expeditiously to remove or disable access to that information, according to Art. 14 (1) of the Directive.

[660] Proposal for the DSA (paras. 40–41); the ECtHR (accepted as a means in cases involving defamation). *See* above in Sect. 3.4.3.2.

[661] For example, the Network Enforcement Act of Germany places obligations on ICTs to remove unlawful speech from their platforms within a specified period of time after receiving notice. *See* Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act) (2017), Sect. 3 (2). Notice may also be given through injections by domestic courts.

[662] De Streel et al. (2018), p. 42.

[663] De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 44.

[664] ibid., p. 10; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 50.

images and copyrighted works.[665] In relation to other forms of content, the adverse effect on human rights, particularly on the freedom of expression, has been observed, given the inability of automated tools to assess nuance and contextual factors.[666] Filtering is often conducted on the basis of keywords or lists of websites, and is frequently void of context, such as cultural considerations, satire, slang, abbreviations and symbols. Filters thus tend to be both under- and overinclusive, censoring also lawful speech.[667] For example, the ECtHR in *Delfi v Estonia* noted the inefficacy of word-based filtering software in detecting hate speech.[668] Simultaneously, filters may disproportionately harm historically marginalised groups.[669] Machine-learning algorithms are instructed by training data, that is, content categorised by humans as harmful will be routed to the algorithm to improve future assessments. There is thus a risk that also AI may reflect and amplify existing social biases, including gender.[670] As noted, this appears to be the case regarding sexist speech, which is common in daily language and may thus be overlooked. However, technology is developing to enhance the accuracy of such tools, including extra-linguistic, knowledge-based features, which consider contextual information and even the background of users, such as their posting history and online identities.[671] Filtering mechanisms more protective of the freedom of expression are also increasingly employed, such as quarantining hate speech.[672] This allows the recipient to deny receipt of flagged material. Nevertheless, issues remain in that filters are easy to circumvent, for example, by using steganography, and does not affect peer-to-peer networks or Usenet newsgroups.

Given the limitations of filters, such must frequently be accompanied by pre- or post-publication human moderation, particularly involving harm requiring contextual assessments, such as hate speech or sexual violence. For example, Facebook has introduced AI tools for the detection of intimate images posted without consent,

[665] For example, Google detection and Microsoft PhotoDNA that find copies of material categorised as unlawful.

[666] CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 2.3.5.

[667] White (2006), p. 107. *See* also Tsesis (2002), para. 76; Meyer (1999), p. 318.

[668] *Delfi v Estonia* (ECtHR), para. 156.

[669] 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, para. 34.

[670] De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 59. Also noted in UN HRC, 'Promotion, Protection and Enjoyment of Human Rights on the Internet: Ways to Bridge the Gender Digital Divide from a Human Rights Perspective' (5 May 2017), para. 41; Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, 'A Union of Equality: Gender Equality Strategy 2020–2025', COM/2020/152.

[671] For example, through neural network embeddings. *See* Badjatiya et al. (2017); Ullmann and Tomalin (2020), p. 73.

[672] Ullmann and Tomalin (2020).

flagged for human review.[673] Additionally, certain media publishers use filters in comments sections, with additional manual reviews of comments linked to articles and materials on sensitive issues.[674] For example, as noted previously, it is more common that articles involving sexual violence or feminist arguments attract irate remarks. Nevertheless, given the global reach of the Internet, moderators may still not be attuned to contextual elements in the particular case, such as cultural nuance. Instances of interpersonal harm, such as harassment, defamation and disclosure of private information may also require awareness of individual circumstances.[675] Additionally, intermediary staff frequently lack sufficient knowledge of domestic laws or international human rights law standards, and how to apply such in complex assessments of content.[676] Monitoring may thus result in overbroad censorship, that is, unlawful restrictions of the freedom of expression, while illicit speech—especially misogynistic language—is overlooked.[677]

Additionally, the approach to notice-and-takedown mechanisms varies across states, including traditional 'notice-and-takedown' systems (the illegal content must be removed), 'notice-and-staydown' systems (the illegal content must be removed and cannot be re-uploaded), or a 'notice-and-notice' system (the hosting provider forwards the notification of infringement to the alleged perpetrator).[678] The necessity of a detailed and harmonised notice-and takedown procedure has thus been called for in the context of EU law, in order to ensure legal certainty.[679] While, for example, the draft DSA obliges hosting service providers to put in place mechanisms of notifications, it does not specify the type of system required.[680] The reason may be that different notice-and-takedown mechanisms could be used depending on the type of liability applicable, that is, whether involving criminal law (e.g. child sexual abuse) or civil law (e.g. defamation). For instance, it has been suggested that a notice-and-notice system would be suitable in relation to copyright infringements, notice-wait-and-takedown involving defamation and notice-and-takedown as well as notice-and-suspension for hate speech.[681] The content provider may, for example,

[673] Facebook, 'Detecting Non-Consensual Intimate Images and Supporting Victims' (15 March 2019) <https://about.fb.com/news/2019/03/detecting-non-consensual-intimate-images/> Accessed 15 March 2022.

[674] *Standard Verlagsgesellschaft MBH v Austria (No. 3)* (ECtHR), para. 9.

[675] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye' (6 April 2018), para. 29.

[676] CoE, 'Recommendation CM/Rec(2018)2 of the Committee of Ministers to Member States on the Roles and Responsibilities of Internet Intermediaries', Principle 2.3.4.

[677] UNESCO, 'Artificial Intelligence and Gender Equality: Key Findings of UNESCO's Global Dialogue' (2020).

[678] Angelopoulos and Smet (2016).

[679] European Parliamentary Research Service, 'Reform of the EU Liability Regime for Online Intermediaries: Background on the Forthcoming Digital Services Act' (May 2020), p. 15.

[680] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC (2020), Art. 14.

[681] Angelopoulos and Smet (2016), p. 267.

need more time to evaluate defamation claims than child abuse images, also allowing room for counter notices. In a study commissioned by the European Parliament, digital platforms in fact indicated a practice of swifter removal of terrorist related content and child abuse images in contrast to other content, not only as a result of the content being categorised as particularly harmful but also in view of more rapid assessments of its illegality, both through machine learning and human moderation.[682] In the same study, numerous digital platforms held that, given the suitability of different mechanisms depending on the harm, a one-size-fits-all approach in terms of moderating measures is not useful, an approach broadly adopted also by scholars.[683] As noted above, assessments of restrictions of online speech by regional human rights law courts and UN treaty bodies frequently balance qualified rights. Such balancing exercises may accordingly consider the particular form of moderation relative to the harm in question.

An additional mechanism of restricting access to online content, be it certain websites or types of material, is the blocking of information.[684] It is generally more efficient to block information than to file claims against authors and publishers. European governments make numerous requests to block or filter content each year, viewed in the biannual Google Transparency Report on removal requests.[685] Such measures raise several legal concerns. The use of blocking technologies by states often violates the freedom of speech of individuals, as they also tend to block lawful speech.[686] Blocking mechanisms are often imprecise and ineffective as they are likely to produce false positives (blocking sites with no prohibited material) or false negatives (sites with prohibited material slip through), and are easy to by-pass, for example, through the creation of mirror websites.[687] Accordingly, it limits both the freedom of expression of the author as well as the purported recipients of information, that is, audience rights. Particularly clear and rigorous safeguards must thus be in place when adopting such measures.[688]

[682] Usually within an hour, as opposed to 24 h for other illegal content. *See* De Streel et al. for the European Parliament, Directorate-General for Internal Policies of the Union (2020), p. 44.

[683] ibid., p. 45.

[684] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 29.

[685] Google Transparency Report: <https://transparencyreport.google.com/?hl=sv> Accessed 8 March 2022.

[686] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 31.

[687] Korff (2014), p. 13.

[688] *See* IACmHR, Office of the Special Rapporteur for Freedom of Expression, 'Freedom of Expression and the Internet' (2013), para. 58; UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 75.

The ECtHR has through case law developed certain safeguards in this respect.[689] For example, in *Cengiz and Others v Turkey*, the extended blocking of YouTube contravened Article 10, affecting the right of individuals to receive and impart information.[690] In this case, there was no provision in the law allowing for a blanket blocking order. In *Ahmet Yildirim v Turkey*, blocking access to an entire site, such as a search engine (Google sites), was considered a breach of Article 10, as it had an impact on the right to receive and impart information, and in effect led to collateral censorship.[691] In *OOO Flavus and Others v Russia*, the blocking of entire domain names rather than specific Uniform Resource Locators (URLs) was disproportionate.[692] Accordingly, any restriction must be limited to what is strictly necessary.[693] Similarly, the UN HRC has held that, with regard to Article 19 of the ICCPR, permissible restrictions of the operation of websites, blogs or any other Internet-based systems should be content specific as generic bans on the operation of certain sites and systems are not compatible with the provision.[694] However, in *Akdeniz v Turkey,* the ECtHR found it acceptable to block certain sites involved in unlawful activities, such as streaming music without respect for copyright legislation. The user was not deemed to have suffered harm in the sense of being a victim.[695]

As for *obligations* to block material, Directive 2011/93/EU on the protection of children from sexual abuse, exploitation and child pornography requires states to remove child abuse images online and encourages the blocking of access to such websites.[696] This approach also appears in soft law.[697] A directive on online public

[689] The collective case law indicates that blocking measures require: (1) a definition of the categories of persons or institutions subject to have their publications blocked; (2) a specification of the categories of blocking orders, for example, whether it involves entire websites or IP addresses etc.; (3) a provision on the territorial scope of the blocking order; (4) a limit on the duration; (5) an indication of the aim of the order; (6) proportionality; (7) necessity; (8) a specification of authorities competent to issue a blocking order; (9) a procedure for reviewing the order; (10) notification of the blocking order and; 11) possibilities of judicial appeal. *See Ahmet Yildirim v Turkey* (ECtHR), Concurring opinion of Judge Albequerque, pp. 27–28.

[690] *Cengiz and Others v Turkey* (ECtHR).

[691] *Ahmet Yildirim v Turkey* (ECtHR), para. 50. Similarly, in a range of cases involving Russia, the ECtHR has found the blocking of entire websites disproportionate, given the collateral effect on lawful speech.

[692] *OOO Flavus and Others v Russia* App nos 12468/15, 23489/15 and 19074/16 (ECtHR, 23 June 2020). Meanwhile, in *Vladimir Kharitonov v Russia* App no 10795/14 (ECtHR, 23 June 2020), the blocking of a website due to sharing the same hosting company and IP-address as another website, while not containing illicit material, was also deemed disproportionate.

[693] *Ahmet Yildirim v Turkey* (ECtHR), para. 47. Prior restraint generates strict scrutiny.

[694] UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para. 43.

[695] *Akdeniz v Turkey* App no 20877/10 (ECtHR, 11 March 2014).

[696] Art. 25 (2) of Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography. Notably, the Budapest Convention does not require filtering or blocking mechanisms.

[697] For example, the UN Special Rapporteur on the Freedom of Expression has held that child abuse images are a clear example where blocking measures are justified, while similarly emphasising necessary safeguards. This includes sufficiently precise national laws and safety measures against

provocation to commit a terrorist offence contains similar obligations.[698] Furthermore, the proposal for an EU directive on violence against women contains an article requiring Member States to take measures to ensure the removal or disabling of access to content categorised as cyber violence, upon a decision by domestic judicial authorities.[699] Such regulations are not in all instances framed as human rights issues, thus not explicitly considering the impact on, for example, the freedom of expression. It is, however, presumed in the directives that blocking must be in accordance with human rights law principles, such as providing information on the reasons for the restrictions.[700] Regardless of the evident legitimacy of the aim, blocking as a measure may still be considered disproportionate, if unsuited to fulfil the aim, excessive in its effect or lacking in procedural safeguards.[701] This may even concern child abuse material.[702] The main focus should thus be on the prosecution of individuals responsible for the production and dissemination of such material rather than blocking measures.[703]

An aspect related to the right to anonymity is the right to be forgotten, which also places duties on online search engines in terms of controlling content. The scope of this right was developed by the ECJ in the *Google Spain* case and, although limited in scope as it pertains to EU Member States, this has been interpreted in a broad manner.[704] The outcome of the case entails that individuals may request Internet search engines to remove links to personal information. It thus extends individual

misuse, such as through an independent regulatory body. *See* UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 32.

[698] Art. 21 (2) of Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism.

[699] Art. 25 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', COM(2022) 105 final (8 March 2022).

[700] This requires transparent procedures and adequate safeguards, for instance, that measures are limited to what is necessary and proportionate. *See* Art. 25 (2) of Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography; Art. 21 (3) of Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism.

[701] Korff (2014), p. 74.

[702] For example, if there is no indication that it leads to a decrease of abuse of children or the sharing of such images. *See* Korff (2014), p. 74.

[703] UNHRC, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011), para. 71.

[704] Case C-131/12, *Google Spain SL v Agencia Espafiola de Proteccion de Datos (AEPD) and Mario Costeja González* (2014), ECLI:EU:C:2014:317. This right draws on Art. 17 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), affirming a right to erasure. The ECJ in Google Spain held that also an EU-based subsidiary of a multinational corporation with headquarters outside of the EU may be subject to EU data protection law, regardless of whether the multinational corporation operates the data processing. The subsidiary is thus subject to the national

control over personal information, in the balance between the freedom of expression and the right to privacy. The ECJ considered that the fundamental rights in Article 7 and Article 8 of the Charter of Fundamental Rights of the European Union, that is, the right to privacy and the right to personal data, 'override, as a rule, not only the economic interest of the operator of the search engine but also the interest of the general public in finding that information upon a search relating to the data subject's name'.[705] In practice, the search engine must assess the request on a case-by-case basis. For example, Google has a webform for requesting the removal of URLs from its index. The company must in turn consider such factors as accuracy, adequacy and relevance (including time passed) of the information.[706]

It should be noted that the ECJ in the case made a distinction between Internet media publishers and operators of search engines. The former are not subject to similar requirements, when performed 'solely for journalistic purposes'.[707] In contrast, search engines make access to information easier for Internet users and may 'play a decisive role in the dissemination of that information', exacerbating the interference.[708] A similar differentiation was made by the ECtHR in *M.L. and W.W. v Germany* involving old articles on a news portal available on the Internet, which identified convicted offenders.[709] While the case concerned access to remedies, connected to the responsibility of news portals to erase information, the ECtHR expounded also on the role of search engines. The ECtHR noted that search engines, even though they do not publish their own information, amplify the effects of interferences. The obligations of search engines *vis-à-vis* the individual who is the subject of the information may thus be different from the website which published the information.[710] In contrast, news websites fill an important role in informing the public, through websites and digital archives, requiring a broader protection of their freedom of expression. Consequently, in relation to such websites, particularly strong reasons must be provided for restricting access to information which the public has a right to receive. [711] In this regard, such factors as whether the information contributed to the public interest, whether the person was well-known, the prior conduct of the person and the content, form and consequences of the publication were assessed. Given that the articles were found on online news portals, they 'were not likely to attract the attention of those Internet users who were not seeking information about the applicants'.[712] Albeit acknowledging the long-lasting

law of the EU member state in which it has its establishment. It entails that the links may still be accessible on Google.com, albeit deleted from European versions. *See* para. 55.

[705] ibid., para. 97.

[706] ibid., para. 92.

[707] ibid., para. 85.

[708] ibid., para. 87.

[709] *M.L. and W.W. v Germany* (ECtHR).

[710] ibid., para. 97.

[711] ibid., para. 102.

[712] ibid., para. 113.

accessibility of information online once published, the option remained for applicants to contact search engines with a view to having the information removed.[713]

The right to be forgotten has been criticised for its limited consideration of the impact on the freedom of expression—in effect restricting the dissemination of information—in addition to such values as democracy.[714] Arguably, since most states have laws prohibiting hate speech, defamation and libel, the right to be forgotten may well be applied mainly to information that is correct, *de facto* entailing a suppression of the truth, contravening one of the main values of protecting the freedom of speech.[715] Furthermore, the standard of what kind of information should be removed is vague, presumably causing search engines to be overcautious.[716] Additionally, if a search engine removes information that does not meet the required standards, no legal repercussions ensue. Again, requests are processed by private companies and the content of human rights are in effect determined by private entities. As such, it can be argued that the right to be forgotten in this form does not in fact ensure individual control over information, as the assessment of removal is delegated to companies.[717]

Nevertheless, from the viewpoint of gender-based offences, the right to be forgotten is a welcome development as it has the potential of reducing reputational, psychological and economic harm resulting, for instance, from a loss of employment opportunities. Beyond the mentioned sources, a right to be forgotten is not explicitly regulated in international human rights law. However, the UN Special Rapporteur on Violence against Women considers the establishment of procedures for the immediate removal of harmful content, the possibility of interim orders and reparations, to be encompassed within the obligation to protect women against online violence, extending also to websites other than search engines.[718] The option of interim orders of removal is also suggested in the proposed EU directive including cyber violence.[719] This can thus arguably be viewed as implicit in obligations to provide effective remedies in relation to gender-based violence.

[713] ibid., para. 114.

[714] Arguably, a sophisticated balancing between such interests was missing in the case, with no discussion of proportionality. *See*, for example, Kohl and Rowland (2017), p. 98.

[715] Integral to the marketplace of ideas theory. *See* Crockett (2016), p. 175.

[716] Crockett (2016), p. 176.

[717] Taddeo and Floridi (2016), p. 1594.

[718] For example, immediate injunctions that prohibit the perpetrator from circulating the material pending a resolution of the legal case, in collaboration with Internet intermediaries. *See* UNHRC, '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), paras. 67, 70, 103.

[719] Art. 25 (2) of European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence' (8 March 2022).

3.4.3.4 Value Embedded Technological Design

Internet intermediaries do not generally perpetrate gender-based offences. At the same time, they are not passive entities. They provide the platforms for abuse and may facilitate or exacerbate offences through their design, content and scope of self-regulation, underpinned by Internet architecture. As noted previously, technological design affects human behaviour, including how we express ourselves, both in terms of social norms and what can be relayed through any particular forum. How ICTs are used and the employed means of monitoring and moderating content may also constrain or safeguard particular values of rights. Although most focus in terms of ICT liability has been placed on the protection of human rights at the content or personal information level, it is increasingly acknowledged that human rights law may influence also product and service development. Whereas Sect. 2.3 considered how certain regulatory tools—such as law and architecture—affect the individual, this part in a general manner addresses how law may or should influence technological design.[720]

Technology is designed based on multiple factors, both formal and informal. Whereas commercial incentives are a driving force, policy decisions are made as to how technology is to be used. Technological development is a social process, shaped *inter alia* by culture, legal frameworks, global financial markets and institutional imperatives.[721] Factors also include social norms, public perception and morality.[722] This is aligned with the cyberfeminist approach discussed previously, which considers that the form, use and content of technological innovation are not inevitable but rather a matter of choice, in tandem with social norms and development. That is, gender roles develop alongside technology while also influencing its design. Law is thus one factor that may encourage, or possibly coerce, technological change. For example, design may be employed as an instrument for implementing such regulatory goals as data protection and privacy.[723] Design may in certain instances be required to meet specific legislative purposes, such as criminal justice. As noted above, this can involve blocking, filtering and surveillance. For example, states may demand that filtering software restricting certain types of content is pre-installed in computers, mobile telephones, switchers and routers. Network managers are able to design equipment that allows for easier interception of data, such as "data packet inspection", to track user activity.[724] Nevertheless, regulatory frameworks are often

[720] *See* discussion in Lessig (1999), p. 512.

[721] Scott et al. (2011), p. 7.

[722] Luger and Golembewksi (2017), p. 300.

[723] With a broad view of regulation—as measures aiming to alter the behavior of others according to defined and agreed upon standards—technological design may itself constitute a system of regulation. *See* Lessig (2006); Luger and Golembewksi (2017), p. 302.

[724] Dunston for BSR, 'Protecting Human Rights in the Digital Age: Understanding Evolving Freedom of Expression and Privacy Risks in the Information and Communications Technology Industry' (2011) <https://www.bsr.org/en/our-insights/report-view/protecting-human-rights-in-the-digital-age> Accessed 30 March 2022, p. 17.

considered as an afterthought by technological designers, rather than a component of the design process.[725] As previously noted, the technical features of the Internet in several instances compel international human rights law to adapt. This is, for example, evident in the assessment of "reasonable expectations of privacy" and the scope of obligations to protect, limited by user anonymity and the ability of intermediaries to control third-party content.

Technological development void of the values entrenched in law carries certain risks. According to Lessig, code—meaning programs and protocols of the Internet—is more effective and precise in constraining individual behaviour than law, given that its enforcement is *ex-ante*, self-executed and generally at a lower cost than traditional law enforcement.[726] Public goods and the moral values of law may in this regard be compromised, with the application of rights being controlled by market forces, potentially enhancing or restraining rights.[727] Code can, for example, embed certain values or make specific values impossible, in effect promoting other aims. For example, as noted by the UN Special Rapporteur, moderation by algorithms reflects the biases of the rule-settlers of online platforms, which are often ethnically and economically homogenous.[728] This includes gender biases in AI.[729] Meanwhile, code is unstable as a constraint, as it is subject to rapid development. The argument is thus that law must remain the primary source of regulation, and to ensure that regulatory objectives are embedded in code—be it voluntarily or through coercion—so that values are not determined by commerce but through the democratic process. Rather than law adapting to Internet architecture, the law should thus provide incentives and/or require businesses to develop technologies that support public policy.[730] Although the discussion on code as a form of regulation has primarily focused on privacy issues and the freedom of expression,[731] there is no inherent limitation as to which aspects may be considered.

Given the transnational nature of the Internet, international human rights law is a reasonable source for identifying international standards to be embedded in code.[732] Not only human rights should be considered, but also the values they aim to protect.[733] There is in fact increasing recognition at the international level that human rights should be integrated into new technology through product

[725] Luger and Golembewksi (2017), p. 307.

[726] Lessig (1999), p. 511.

[727] Lessig (1999), p. 523; Lessig (2006), p. 191. *See* also Spinello (2016), p. 5.

[728] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 84.

[729] Femm committee, European Parliament, 'Cyber violence and hate speech online against women' (2018), p. 18.

[730] Reindenberg (2005), p. 1969.

[731] Luger and Golembewksi (2017), p. 303; Balkin (2004), p. 49.

[732] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 84.

[733] Lessig (2006), p. 165.

development and design by default, for example, to address hate speech.[734] As mentioned previously, the UN Guiding Principles require businesses to conduct human rights impact assessments, that is, to consider how products will be used by customers, in order to develop strategies to mitigate the risk of misuse. This particularly includes the impact on the freedom of expression and the right to privacy of users.[735] It is also proposed in the DSA of the EU, *vis-à-vis* VLOPs. As viewed in the section on corporate codes of conduct, there is also an increased collaboration between ICTs and states to create and adapt products and technological infrastructure in alignment with human rights law norms.

However, in requiring the adaptation of Internet architecture through law, the manner in which the Internet affects the values of human rights must be considered. As viewed, the Internet exacerbates clashes between certain rights and may affect how values are ensured or constrained. Concretely, if a feature interferes with a particular end value, law may regulate to change the architectural feature or accept that the value cannot be ensured to the same extent online.[736] For example, John Balkin argues that free speech values must be integrated into ICTs.[737] Meanwhile, various features of the Internet reflect and enhance the value of individual autonomy of the freedom of expression. In turn, the value of individual autonomy is in many respects aligned with the libertarian approach, allowing for a wide scope of individual choice among different values in order not to constrain liberty, which is not reflected in international human rights law.[738] Human rights law may thus either influence the adaptation of design in line with a more restrictive freedom of expression, in order to ensure an online/offline coherence, or accept the difference of values online.

As mentioned, gender equality is a core value in international human rights law. Whereas the focus of liberal feminism in the 1970s was on the inclusion of women in employment and education in science as a means of ensuring gender equal Internet access and design,[739] the development of gender-sensitive technological features is increasing. In this regard, human rights impact assessments may be employed to examine the use and social impact of products from a gender perspective when

[734] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), paras. 42, 44; UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. David Kaye' (30 March 2017), para. 59; European Commission, 'White paper on Artificial Intelligence – A European approach to excellence and trust', COM(2020) 65 final: <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf>Accessed 9 March 2022.

[735] *See* UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. David Kaye' (30 March 2017), para. 53, for examples of what this may entail in the context of the Internet.

[736] Lessig (1999), p. 505.

[737] Balkin (2004), p. 49.

[738] Post (2000), p. 1440.

[739] Wajcman (2004), p. 14.

developing new technologies, that is, a gender mainstreaming of Internet architecture. This has been noted as an objective by the UN Human Rights Council, the UN Special Rapporteur on the Freedom of Expression and the EU.[740] What gender-sensitivity entails is, however, contentious and varies according to an array of feminist theories *vis-à-vis* different types of content, for instance, in relation to pornography. Although there are common aims, such as preventing gender-based violence, the suggested means to achieve gender equality thus differ. For instance, from the perspective of cultural and dominance feminist theories, female Internet users have different experiences—including ways of communicating—and thus needs, to which Internet design must be adapted.

Bearing in mind gender in technological development may range from considering female aesthetic in the design of websites to using code to prevent and report gender-based harm. As discussed previously, the increased use of algorithms detecting offences may aid the prevention of offences. New developments include technology identifying images of a sexual nature that may have been shared without the subject's consent, implemented by certain social media companies. As noted, machine learning and AI will be used to proactively detect near-nude images or videos in order to remove them before they are reported.[741] Other possible measures include automatic hate speech detection on social media through algorithms. However, as noted above, lexical detection methods tend to be both over- and underinclusive and frequently display a gender bias. Studies indicate that especially sexist speech, as opposed to racist hate speech, is less likely to be detected.[742]

As the responsibility for content control and reporting of violations mainly lies on the victim or other users, increasing focus is placed on developing easier navigation of websites and access to safety tools.[743] Users often have difficulties understanding privacy settings and such are generally 'public by default, private through effort'.[744] Measures have thus been taken by certain platforms to increase user control over privacy, not only through transparency, but through interface.[745] Additionally,

[740] UN HRC, 'Resolution adopted by the Human Rights Council on 5 July 2018: Accelerating Efforts to Eliminate Violence against Women and Girls: Preventing and Responding to Violence against Women and Girls in Digital Contexts' (17 July 2018) UN Doc. A/HRC/RES/38/5, para. 10 (d); UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 98; European Commission, 'White paper on Artificial Intelligence – A European approach to excellence and trust'.

[741] Facebook, 'Detecting Non-Consensual Intimate Images and Supporting Victims' (15 March 2019) <https://about.fb.com/news/2019/03/detecting-non-consensual-intimate-images/> Accessed 15 March 2022. *See* also "image-shield" developed by the World Wide Web Foundation, Tech Policy Design Lab <https://techlab.webfoundation.org/ogbv/prototypes#image-shield> Accessed 15 March 2022.

[742] Davidson et al. (2017).

[743] *See*, for example, World Wide Web Foundation, Tech Policy Design Lab: Online Gender-Based Violence and Abuse <https://techlab.webfoundation.org/ogbv/prototypes#reporteroo> Accessed 8 March 2022.

[744] Boyd (2010). *See* also Tene and Polonetsky (2014), p. 83.

[745] Murphy (2016), p. 7.

means for users to control responses to posts have been developed. This includes settings where users can block or mute accounts flagged as inauthentic or that have previously transgressed terms of service, or by choosing individual keyword filters for comments, in addition to the website's own filters. Through new forms of notification settings, users can turn off comments or downloads in cases where content has unexpectedly gone viral.[746] Furthermore, whereas most platforms use notice-and-takedown mechanisms—construed as an obligation in relation to certain forms of content—it can be enhanced through "urgent reaction" reporting systems in response to harmful conduct on social media websites, for example, involving live feed broadcasting illegal acts.[747] Reporting systems that give short explanations of categories of abuse through hover buttons and allow for the adding of contextual information to reports—including cultural and linguistic nuance—have also been designed.[748]

Nevertheless, adaptation of Internet architecture will not be made of all features exacerbating gendered harm, be it due to technical constraints or by value preference. For example, from the perspective of modifying features undermining the regulation of individual offenders, the blanket removal of user anonymity would be the primary option. However, user anonymity is not only an integral feature of the Internet but also considered a human right. Similarly, technological means of effectively monitoring speech pre-publication may prevent harm but also undermine the freedom of expression. A coherent approach to how rights apply online and which values are to be embedded in Internet architecture—in consideration of technological constraints—is thus necessary.

3.4.4 Conclusion

The ambiguity in international human rights law on the scope of liability for Internet intermediaries and online media publishers is related to the nature of public international law—regulating the acts and omissions of states—in addition to the continuous development of Internet architecture. For example, since the promulgation of the e-Commerce Directive, the social web has expanded exponentially, further increasing the complexity in categorising online platforms and thus the scope of liability. Furthermore, it is apparent that the matter to an extent is approached differently in international human rights law, as opposed to from an economic free movement perspective. This entails that the balancing between individual and

[746] *See*, for example, World Wide Web Foundation, Tech Policy Design Lab <https://techlab.webfoundation.org/ogbv/prototypes#calm-the-crowd> Accessed 8 March 2022.

[747] BuzzFeed, 'When rape is broadcast live on the Internet' (20 April 2016) <https://www.buzzfeed.com/rossalynwarren/when-rape-is-broadcast-live-on-the-internet> Accessed 15 March 2022.

[748] *See*, for example, World Wide Web Foundation, Tech Policy Design Lab <https://techlab.webfoundation.org/ogbv/prototypes#calm-the-crowd> Accessed 8 March 2022.

corporate interests to an extent may vary depending on the area of law. Despite such fundamental differences, the tentative development on this issue is largely cohesive within the context of European institutions.

In terms of direct liability for publishers, this may arise where operators publish their own content, for example, involving hate speech, obscene pornography or defamation. As will be discussed in the subsequent chapter, the state is generally considered to be within its margin of appreciation to restrict such speech by publishers. Solely under limited circumstances—for example, concerning racist and inciting speech in the ICERD—is there an obligation. As for secondary liability, the ECtHR has in certain instances held that states are either entitled or obliged to adopt a legal regime for holding Internet intermediaries and media publishers liable. The cases have mainly involved the question of whether states may legitimately restrict the operation of various websites. However, the Court has also noted obligations in terms of protecting the privacy rights of individuals (*Delfi v Estonia*).

Liability does not extend to all forms of unlawful speech nor all intermediaries. A distinction is primarily made on the basis of the *type* of Internet intermediary and the *content* of speech. Although the Court has referenced the e-Commerce Directive in its case law, it has not explicitly and independently categorised digital platforms in that light, but rather accepted domestic assessments. Nevertheless, major media enterprises that provide their own content and editorial control over content posted, such as in *Delfi v Estonia*, have not been approached as intermediaries in the manner of benefiting from the exemption of liability, which is aligned with the e-Commerce Directive. A similar approach was adopted in relation to news portals in *MTE and Index v Hungary.* These were instead considered media publishers. Nevertheless, given the characteristics of the Internet, they were treated differently from print publishers in terms of extending liability for third-party content under certain circumstances. Although the Court in the other cases also did not explicitly categorise the corporations as intermediaries according to the Directive, its approach appears to be that of considering blog service providers and search engines—such as Google—as hosts which, according to the Directive, would remove exemption of liability upon knowledge of unlawful content. According to the ECJ, these may fulfil the role of hosting services, depending on the characteristics of the corporation in question.[749] Meanwhile, third-party liability for public figures managing social media pages was addressed in *Sanchez v France* and *Jezior v Poland*, where the applicant in the former case was categorised as a publisher and the latter as a host, in view of such contextual elements as the aim and readership of the posts.

The ECtHR has, however, in several regards departed from the e-Commerce Directive, indicating a pragmatic approach to liability, considering the financial and practical capabilities of intermediaries in controlling and removing harmful material. The scope of obligations has to a degree been considered relative not only to the passive or active role of the platform but also in relation to the business model, the

[749] *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* (2012), para. 27.

financial capacity and size of intermediaries and media publishers as well as the volume of their user base. As such, the Court has placed a greater degree of responsibility on major, commercial platforms. Such a differentiated approach is increasingly adopted at the international level, evident in the Audiovisual and Media Service Directive, the DSA, CoE Recommendations and in the view of various UN agencies.

Furthermore, the content of the material affects state obligations. EU law categorises hate speech, child abuse images, copyright infringements and terrorist content as "illegal", with stricter obligations for intermediaries. The proposed directive on violence against women extends this to rape and certain forms of cyber violence. In contrast, harmful content generates more limited obligations for VLOPs under the proposed DSA. Whereas this distinction in part stems from a lack of agreement among states in regulating certain forms of online content, it also reflects a hierarchy of the perceived severity of harm. This distinction is also increasingly employed at the domestic level.

Meanwhile, obligations for media publishers to ensure secondary liability has been affirmed by the ECtHR in cases of hate speech or direct threats of violence against individuals, as opposed to defamation. Although the Court has indicated that also defamation is considered "clearly unlawful" speech, a distinction has been made in practice, solely extending such liability to hate speech. According to the Court in *Delfi v Estonia,* notice-and-takedown systems are not adequate in relation to hate speech, which implies an obligation of active pre-publication monitoring or filtering. In contrast, in *Pihl v Sweden* and *Payam Tamiz v the United Kingdom*, the speech was considered vulgar, offensive and defamatory—but not hate speech—with individuals consequently acquiring more limited protection. In *Høiness v Norway*, the Court did not consider it necessary to analyse the comments as instances of sexual harassment, as they clearly did not constitute hate speech nor incitement to violence. Notably, the context of the Internet affects the evaluation of the egregious nature of speech, where a higher threshold of tolerance is presumed by *inter alia* the ECtHR. That is, what is "clearly unlawful" may pertain to comments of a graver nature than IRL.

As such, unless sexist speech is categorised as hate speech, and certain forms of harassment considered threats of violence, no *explicit* obligation has been placed on states to ensure secondary liability for intermediaries or media publishers in relation to gender-based harm in the case law of the ECtHR. Nevertheless, speech outside the scope of "clearly unlawful" may still generate state obligations to regulate intermediaries, as an aspect of obligations to protect individuals. In the ECtHR cases on defamation and harassment, the Court accepted the notice-and-takedown processes employed by intermediaries as suitable means of addressing these forms of harm. There is also room for developing a gender-sensitive approach to what is considered "clearly unlawful" by the ECtHR. For example, it is apparent that states incur obligations to criminalise sexual violence—whether speech-based or physical—which may encompass sexual harassment. However, such an approach was not accepted by the ECtHR in *Høiness v Norway*. In the context of the EU, certain forms of cyber violence may in the future be categorised

as "illegal", if the proposal for a directive results in regulation. Nevertheless, various forms of gender-based harm and materials online will still most likely be considered "harmful" content. This includes harmful pornography, interpersonal harassment and forms of sexual violence not encompassed by the narrow definition of rape in the proposal.

The reason for this ranking of content appears to be in part ideological, related to the hierarchy of protected speech, and in part practical. International human rights law clearly establishes that hate speech does not benefit from protection of the freedom of expression. It appears that pragmatism is also a prominent consideration, with the ECtHR indicating that intermediaries do not have to perform an advanced legal assessment in relation to hate speech. In contrast, it may be difficult for corporations to assess what constitutes defamation, harassment or the disclosure of private information. In view of this practical approach, it should, however, be noted that legal assessments of what constitutes hate speech cannot be understood as straightforward *per se*, unless involving overtly degrading speech. As noted, hate speech is contextually evaluated, bearing in mind the cultural, demographic and historical background of the speaker and audience. This creates particular complexities on the Internet, especially on websites attracting a global audience. Meanwhile, sexually harassing and degrading speech may, arguably, require less advanced legal assessments, as such forms of harm to a degree involve objective standards. Additionally, as noted previously, evaluations of the intent and effect of speech are more challenging on the Internet in general, given the lack of social cues and with indicia including multimodal expressions, the use of emoticons, steganography and the jargon of particular platforms.

The social costs of precautionary measures are an additional consideration. This comports with the approach of the e-Commerce Directive, that the extent of liability is relative not only to the gravity of harm caused by the illegal activity but also the harm caused by incidentally restricting lawful speech.[750] That is, the broader effects of restricting speech on a public forum such as the Internet requires a balancing act, commonly to the benefit of a generous freedom of expression, which may be detrimental to addressing gender-based online harm. The ECtHR has also affirmed that states have a wide margin of appreciation in regulating the liability of intermediaries given the important role they perform in facilitating access to information and debate. As liability in this fair balancing approach is relative to the harm, it is imperative that the assessment of harm is gender-sensitive.

This delineation of different forms of content does not preclude states from extending liability for such acts and materials at the domestic level. For example, it has been noted that so-called revenge pornography is increasingly illegal in EU Member States. However, the lack of international regulation imparts the view that gender-based violations are less harmful, at both the individual and social level. From a practical perspective, the lack of harmonisation undermines the enforcement

[750] EU Parliament, Directorate General for Internal Policies, 'Providers Liability: From the eCommerce Directive to the Future: In-depth analysis for the IMCO Committee', p. 30.

of domestic laws and clearer guidelines for intermediaries in terms of monitoring and moderation, thus affecting the ability to effectively regulate this sphere. As many intermediaries are located in the US—with a broad freedom of expression—intermediary liability is in practice subject to a lower threshold.[751]

The scope of acceptable intermediary liability has additionally been considered relative to the size of the audience, which recognises that a wide user base limits the possibilities of actively monitoring content. This has also been an aspect in the assessment of harm to the victim. In *Tamiz v the United Kingdom* and *Pihl v Sweden*, the ECtHR considered whether the blog/website was well-known and attracted a broad readership, that is, the effect of the defamatory statements. However, the architecture of the Internet should be borne in mind when assessing harm, considering the ease with which obscure websites can be accessed through search engines. As websites are not necessarily listed in chronological order, outdated, defamatory information can easily be retrieved. At the same time, as addressed by the ECtHR in *M.L. and W.W. v Germany*, search engines amplify access to and dissemination of information, which does not affect the liability of news portals themselves.

In conclusion, a notice-and-takedown system will most likely be considered the most appropriate form of monitoring and moderation of content, in relation to a range of gender-based harm, albeit not yet explicitly construed as an obligation. However, as discussed, given the difference in nature of these forms of harm and available tools for detection and removal, it may involve variations of the notice-and-takedown regime. For example, notice-wait-and-takedown may be appropriate in relation to defamation and disclosure of personal information, whereas image-based sexual abuse may be detected through filters and human moderation. At the same time, a notice-and-takedown approach is a reactive policy. It does not prevent the dissemination from occurring, thus addressing the symptom and not the cause. Also, such a mechanism does not assist individuals that are not immediately aware they have been maligned. Such means must thus be accompanied by additional and parallel measures, including the adaptation of ICT architecture and design to international human rights law standards, including the protection of individuals against gender-based violations. More advanced automatic tools detecting offences, attuned to contextual nuance, and settings allowing users variations in the level of control of published materials and responses will aid efforts of prevention. Nevertheless, beyond the development of notification systems, and certain due diligence requirements for VLOPs in the DSA, ICTs are not under direct obligations to consider human rights law at the design stage.

On a general note, the ECtHR in several cases held that although user anonymity is an obstacle in holding individual perpetrators accountable, it is still the primary venue for liability. It also clear that the identification of perpetrators is part of state obligations to protect individuals against certain grave offences. It should thus be emphasised that state obligations to adopt effective criminal or civil laws *vis-à-vis*

[751] CoE, 'Seminar Combating Sexist Hate Speech: Report', 10–12 February, EYC, Strasbourg, p. 29.

individual perpetrators remain the same online and distinctions in terms of content, such as stronger protection against hate speech, are not indicative of the approach to obligations against the authors of comments. For example, online sexual harassment generates similar obligations of regulation as offline. Nonetheless, as noted by the domestic courts in *Payam Tamiz v the United Kingdom* and *Høiness v Norway*, the difficulty in verifying the identity of the authors of comments entails that individual liability is more theoretical than real. With limited liability for intermediaries, practical possibilities for victims of online gender-based harm to access remedies are thus minor.

References

Akdeniz Y (1997) The regulation of pornography and child pornography on the internet. JILT 1997: 1–38

Alexander L, Horton P (1983–1984) Review essay: the impossibility of a free speech principle. Northwest Univ Law Rev 78:1319–1357

Allen A (1987) Taking liberties: privacy, private choice, and social contract theory. Univ Cincinnati Law Rev 56:461–491

Allen A (1999) Coercing privacy. William Mary Law Rev 40:723–757

Angelopoulos C, Smet S (2016) Notice-and-fair-balance: how to reach a compromise between fundamental rights in European intermediary liability. J Media Law 8:266–301

Asquith N (2007) Speech act theory, maledictive force and vilification in Australia. In: Ensor J et al (eds) New talents 21C: other contact zones. Network Books, Perth, pp 179–188

Austin JL (1975) How to do things with words, 2nd edn. Harvard University Press, Cambridge

Baber HE (1987) How bad is rape? Hypatia 2:125–138

Baines B (2009) Contextualism, feminism, and a Canadian woman judge. Fem Leg Stud 17:27–42

Balkin JM (2004) Digital speech and democratic culture: a theory of freedom of expression for the information society. N Y Univ Law Rev 79:1–55

Balkin JM (2014) Old-school/new-school speech regulation. Harv Law Rev 127:2296–2342

Barak A (2005) Sexual harassment on the internet. Soc Sci Comput Rev 23:77–92

Barbieri F et al (2016) How cosmopolitan are emojis? Exploring emojis usage and meaning over different languages with distributional semantics. In: Hanjalic A et al (eds) Proceedings of the 2016 ACM conference on multimedia conference. https://dl.acm.org/doi/10.1145/2964284.2967278. Accessed 28 Jan 2022

Barendt E (2005) Freedom of speech, 2nd edn. Oxford University Press, New York

Barendt E (2016) Anonymous speech: literature, law and politics. Hart, Oxford

Badjatiya P et al (2017) Deep learning for hate speech detection in tweets. Proceedings of ACM WWW'17. https://www.researchgate.net/publication/317300025_Deep_Learning_for_Hate_Speech_Detection_in_Tweets. Accessed 18 Feb 2022

Bell MC (2021) John Stuart Mill's harm principle and free speech: expanding the notion of harm. Utilitas 33:162–179

Benedek W, Kettemann MC (2013) Freedom of expression and the internet. Council of Europe Publishing, Strasbourg. https://bookcoeint/en/human-rights-and-democracy/5810-freedom-of-expression-and-the-internethtml. Accessed 28 Jan 2022

Bennett JT (2016) The harm in hate speech: a critique of the empirical and legal bases of hate speech regulation. Hastings Const Law Q 43:445–536

Berger G (2017) The universal norm of freedom of expression - towards an unfragmented internet. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 27–38

Biber J et al (2002) Sexual harassment in online communications: effects of gender and discourse medium. Cyberpsychol Behav 5:33–42
Bjarnadottir MR (2016) Does the internet limit human rights protection? The case of revenge porn. JIPITEC 7:204–215
Blume P (2010) Data protection and privacy – basic concepts in a changing world. Scand Stud Law 56:152–164
Bonnici JPM, de vey Mestdagh CNJ (2005) Right vision, wrong expectations: the European Union and self-regulation of harmful internet content. Inf Commun Technol Law 14:133–149
Boyd D (2010) Making sense of privacy and publicity. SXSW keynote speech. https://www.danah.org/papers/talks/2010/SXSW2010.html. Accessed 3 June 2021
Brandeis L, Warren S (1890) The right to privacy. Harv Law Rev 4:193–220
Brax D (2016) Hate speech and the distribution of the costs and benefits of freedom of speech. In: Edström M et al (eds) Blurring the lines: market-driven and democracy-driven freedom of expression. Nordicom, Gothenburg, pp 185–192
Brems E (2017) Conclusion—conflicting views on conflicting rights. In: Smet S, Brems E (eds) When human rights clash at the European Court of Human Rights: conflict or harmony? Oxford University Press, Oxford, pp 242–250
Brennan F (2009) Legislating against internet race hate. Inf Commun Technol Law 18:123–153
Brison SJ (1998) Speech, harm, and the mind-body problem in first amendment jurisprudence. Leg Theory 4:39–61
Broeders D, Taylor L (2017) Does great power come with great responsibility? The need to talk about political responsibility. In: Taddeo M, Floridi L (eds) The responsibilities of online service providers. Springer, Cham, pp 315–323
Brown A (2015) Hate speech law: a philosophical examination. Routledge, New York
Brüggemeier G (2010) Protection of personality rights in the law of delict/torts in Europe: mapping out paradigms. In: Brüggemeier G et al (eds) Personality rights in European torts law. Cambridge University Press, Cambridge, pp 5–37
Butler J (1997) Excitable speech: a politics of the performative. Routledge, New York
Cariolou L (2017) Circumnavigating the conflict between the right to reputation and the right to freedom of expression. In: Smet S, Brems E (eds) When human rights clash at the European Court of Human Rights: conflict or harmony? Oxford University Press, Oxford, pp 171–191
Carr C et al (2012) Speech acts within Facebook status messages. J Lang Soc Psychol 31:176–196
Carr N (2010) The shallows: what the internet is doing to our brains. W. W. Norton, New York
Chaffin S (2008) The new playground bullies of cyberspace: online peer sexual harassment. Howard Law J 51:773–818
Chander A (2010) Youthful indiscretion in an internet age. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 124–139
Chang C-H (2015) New technology, new information privacy: social-value-oriented information privacy theory. N T U Law Rev 10:127–174
Citron DK (2009a) Cyber civil rights. Boston Univ Law Rev 89:61–125
Citron DK (2009b) Law's expressive value in combating cyber gender harassment. Mich Law Rev 108:373–415
Citron DK (2010) Civil rights in our information age. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 31–49
Citron DK (2014) Hate crimes in cyberspace. Harvard University Press, Cambridge
Citron DK, Franks MA (2014) Criminalizing revenge porn. Wake Forest Law Rev 49:345–391
Citron DK, Franks MA (2020) The internet as a speech machine and other myths confounding section 230 reform. Univ Chic Leg Forum 2020:45–75
Citron DK, Penney JW (2019) When law frees us to speak. Fordham Law Rev 87:2317–2335
Conaghan J (1996) Gendered harms and the law of tort: remedying (sexual) harassment. Oxford J Leg Stud 16:407–431

Crockett M (2016) The internet (never) forgets. SMU Sci Technol Law Rev 19:151–181
Davidson T et al (2017) Automated hate speech detection and the problem of offensive language. ICWSM 2017. https://arxiv.org/abs/1703.04009. Accessed 10 Feb 2022
De Streel A et al (2018) Liability of online hosting platforms – should exceptionalism end? CERRE, Brussels. https://cerre.eu/wp-content/uploads/2020/06/180912_CERRE_LiabilityPlatforms_Final_0.pdfp. Accessed 17 Feb 2022
De Streel A et al for the European Parliament, Directorate-General for Internal Policies of the Union (2020) Online platforms' moderation of illegal content online: laws, practices and options for reform. https://op.europa.eu/en/publication-detail/-/publication/cd388309-cc89-11ea-adf7-01aa75ed71a1. Accessed 11 Feb 2022
Diggelmann O, Cleis MN (2014) How the right to privacy became a human right. Hum Rights Law Rev 14:441–458
Dodge A (2016) Digitizing rape culture: online sexual violence and the power of the digital photograph. Crime Media Cult 12:65–82
Dworkin R (1977) Taking rights seriously. Harvard University Press, Cambridge
Fascendini F, Fialová K (2011) Voices from digital spaces: technology related violence against women. Association for progressive communications. https://www.apc.org/sites/default/files/APCWNSP_MDG3advocacypaper_full_2011_EN_0.pdf. Accessed 28 Jan 2022
Feinberg J (1984) Harm to others: the moral limits of criminal law. Oxford University Press, Oxford
Feinberg J (1985) Offense to others: the moral limits of criminal law. Oxford University Press, Oxford
Fentonmiller KR (1994) Verbal sexual harassment as equality-depriving conduct. Univ Mich J Law Reform 27:565–611
Fenwick H, Philipson G (2003) Text, cases and materials on public law & human rights, 2nd edn. Cavendish Publishing, London
Foucault M (1995) Discipline and punish: the birth of the prison. Vintage Books, New York
Ford SM (2011) Reconceptualizing the public/private distinction in the age of information technology. Inf Commun Soc 14:550–567
Ford TE et al (2015) Sexist humor as a trigger of state self-objectification in women. Humor 28: 253–269
Fox J et al (2015) Perpetuating online sexism offline: anonymity, interactivity, and the effects of sexist hashtags on social media. Comput Hum Behav 52:436–442
Fredrickson BL, Roberts T-A (1997) Objectification theory: toward understanding women's lived experiences and mental health risks. Psychol Women Q 21:173–206
Gagliardone I et al for UNESCO (2015) Countering online hate speech. https://unesdoc.unesco.org/ark:/48223/pf0000233231. Accessed 11 Feb 2022
Geneus J (2018) Emoji: the caricatured lawsuit. Colo Technol Law J 16:431–454
Gilden A (2016) Punishing sexual fantasy. William Mary Law Rev 58:419–491
Gilligan C (1982) In a different voice: psychological theory and women's development. Harvard University Press, Cambridge
Gross H (1979) A theory of criminal justice. Oxford University Press, Oxford
Grundlingh L (2018) Memes as speech acts. Soc Semiot 28:147–168
Habermas J (1996) Between facts and norms. Polity Press, Cambridge
Halder D, Jaishankar K (2011) Cyber gender harassment and secondary victimization: a comparative analysis of the United States, the UK, and India. Vict Offenders 6:386–398
Henry N, Powell A (2015) Beyond the "sext": technology-facilitated sexual violence and harassment against adult women. Aust N Z J Criminol 48:104–118
Herring S (1996) Gender and democracy in computer-mediated communication. In: Kling R (ed) Computerization and controversy: value conflicts and social choices, 2nd edn. Academic, San Diego
Herring S, Dainas A (2020) Gender and age influences on interpretation of emoji functions. ACM Trans Soc Comput 3:1–26

Hill R (2015) Cyber-misogony: should 'revenge porn' be regulated in Scotland, and if so, how? SCRIPTed 12:117–140
Hughes K (2012) A behavioural understanding of privacy and its implications for privacy law. Mod Law Rev 75:806–836
Jacobson D (2000) Mill on liberty, speech, and the free society. Philos Public Aff 29:276–309
Jay T, Janschewitz K (2008) The pragmatics of swearing. J Politeness Res Lang Behav Cult 4:267–288
Jay T (2009) Do offensive words harm people? Psychol Public Policy Law 15:81–101
Joppke C (2013) Double standards? Veils and crucifixes in the European legal order. Eur J Sociol 54:97–123
Karanasiou AP (2016). Law encoded: towards a free speech policy model based on decentralized architectures. First Monday 21. https://doi.org/10.5210/fm.v21i12.7118. Accessed 16 Feb 2022
Kernohan A (1993) Accumulative harms and the interpretation of the harm principle. Soc Theory Pract 19:51–72
Kofman E (2003) Rights and citizenship. In: Agnew J et al (eds) A companion to political geography. Blackwell, Malden, pp 393–407
Kohl U, Fox C (2017) Introduction: internet governance and the resilience of the nation state. In: Kohl U (ed) The net and the nation state: multidisciplinary perspectives on internet governance. Cambridge University Press, Cambridge, pp 1–23
Kohl U, Rowland D (2017) Censorship and cyberborders through EU data protection law. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 93–109
Koops B-J, Galic M (2017) Conceptualizing space and place: lessons from geography for the debate on privacy in public. In: Timan T et al (eds) Privacy in public space: conceptual and regulatory challenges. Edward Elgar, Cheltenham, pp 19–46
Korff D (2014) The rule of law on the internet and in the wider digital world. Council of Europe. https://rmcoeint/16806da51c. Accessed 9 Feb 2022
Laid J et al (2021) Relationships between coerced sexting and differentiation of self: an exploration of protective factors. Theol Sex 2:468–482
Laidlaw E (2015) Regulating speech in cyberspace: gatekeepers, human rights and corporate responsibility. Cambridge University Press, Cambridge
Land M (2019) Regulating private harms online: regulation under human rights law. In: Jorgensen RF (ed) Human rights in the age of platforms. The MIT Press, Cambridge, pp 285–316
Langlois G, Slane A (2017) Economies of reputation: the case of revenge porn. Commun Crit Cult Stud 14:1–19
Langton R (1993) Speech acts and unspeakable acts. Philos Public Aff 22:293–330
Langton R (2011) Symposium on Langton's sexual solipsism. Jurisprudence 2:379–440
Leiter B (2010) Cleaning cyber-cesspools: Google and free speech. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 155–173
Lemmens K (2003) Protection of privacy between a rights-based and a freedom-based approach: what the Swiss example can teach us. Maastricht J Eur Comp Law 10:381–403
Lessig L (1999) The law of the horse: what cyberlaw might teach. Harv Law Rev 113:501–549
Lessig L (2006) Code and other laws of cyberspace: version 2.0. Basic Books, New York
Levmore S (2010) The internet's anonymity problem. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 50–67
Levmore S, Nussbaum MC (2010) Introduction. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 1–14
Levy K (2015) Intimate surveillance. Idaho Law Rev 51:679–693
Lidsky L, Cotter T (2007) Authorship, audiences, and anonymous speech. Notre Dame Law Rev 82:1537–1603
Lidsky L (2018) #I U: considering the context of online threats. Calif Law Rev 106:1885–1928

Lievens E (2010) Protecting children in the digital era: the use of alternative regulatory instruments. Martinus Nijhoff, Leiden
Luger E, Golembewksi M (2017) Towards fostering compliance by design: drawing designers into the regulatory frame. In: Taddeo M, Floridi L (eds) The responsibilities of online service providers. Springer, Cham, pp 295–311
MacKinnon C (1987) Feminism unmodified: discourses on life and law. Harvard University Press, Cambridge
MacKinnon C (1989) Toward a feminist theory of the state. Harvard University Press, Cambridge
MacKinnon C (1993) Only words. Harvard University Press, Cambridge
MacKinnon C (2007) Women's lives, men's laws. Harvard University Press, Cambridge
MacKinnon R et al for UNESCO (2014) Fostering freedom online: the role of internet intermediaries. https://unesdoc.unesco.org/ark:/48223/pf0000231162. Accessed 18 June 2020
Maris C (2013) Pornography is going on-line: the harm principle in Dutch law. Law Democr Dev 17:1–23
Massaro T (1991) Equality and freedom of expression: the hate speech dilemma. William Mary Law Rev 32:211–265
Matsuda M (1989) Public response to racist speech: considering the victim's story. Mich Law Rev 87:2320–2381
McCloskey HJ (1998) Mill's liberalism. In: Smith GW (ed) John Stuart Mill's social and political thought: critical assessments, Freedom, vol II. Routledge, London, pp 49–65
McGowan MK (2003) Conversational exercitives and the force of pornography. Philos Public Aff 31:155–189
McGrogan D (2016) The problem of causality in international human rights law. Int Comp Law Q 65:615–644
McNealy J (2012) The privacy implications of digital preservation: social media archives and the social networks theory of privacy. Elon Law Rev 3:133–160
Megarry J (2014) Online incivility or sexual harassment? Conceptualizing women's experiences in the digital age. Women's Stud Int Forum 47:46–55
Meiklejohn A (1948) Free speech and its relation to self-government. Harper & Brothers, New York
Meiklejohn A (1961) The first amendment is an absolute. Sup Ct Rev 1961:245–266
Meyers DT (1995) Rights in collision: a non-punitive, compensatory remedy for abusive speech. Law Philos 14:203–243
Meyer C (1999) Women and the internet. Tex J Women Law 8:305–324
Mikkola M (2019) Pornography: a philosophical introduction. Oxford University Press, Oxford
Mill JS (1859) On liberty, 2nd edn. John W. Parker and Son, London
Mitchell WJ (1995) City of bits: space, place, and the infobahn. The MIT Press, Cambridge
Moon R (1985) The scope of freedom of expression. Osgoode Hall Law J 23:331–357
Moore B (1984) Privacy: studies in social and cultural history. Routledge, New York
Munro V (2007) Dev'l-in disguise? Harm, privacy and the Sexual Offences Act 2003. In: Munro V, Stychin C (eds) Sexuality and the law—feminist engagements. Routledge, London, pp 1–18
Murphy MH (2016) Technological solutions to privacy questions: what is the role of law? Inf Commun Technol Law 25:4–31
Murray A (2016) Information technology law: the law and society, 3rd edn. Oxford University Press, Oxford
Netanel NW (2000) Cyberspace self-governance: a skeptical view from liberal democratic theory. Calif Law Rev 88:395–498
Neu J (2008) Sticks and stones: the philosophy of insults. Oxford University Press, Oxford
Nielsen LB (2002) Subtle, pervasive, harmful: racist and sexist remarks in public as hate speech. J Soc Issues 58:265–280
Nussbaum M (1997) The feminist critique of liberalism. The Lindley Lecture. The University of Kansas
Nussbaum M (2004) Hiding from humanity: disgust, shame and the law. Princeton University Press, Princeton

Nussbaum M (2010) Objectification and internet misogyny. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 68–90
Nussberger A (2020) The European Court of Human Rights. Oxford University Press, Oxford
O'Connell R (2020) Law, democracy and the European Court of Human Rights. Cambridge University Press, Cambridge
Panichas GE (2014) An intrusion theory of privacy. Res Publica 20:145–161
Pavan E (2017) Internet intermediaries and online gender-based violence. In: Segrave M, Vitis L (eds) Gender, technology and violence. Routledge, London, pp 62–78
Pejchal V (2020) Hate speech regulation in democracy. Routledge, London
Persak N (2007) Criminalising harmful conduct: the harm principle, its limits and continental counterparts. Springer, New York
Post DG (2000) What Larry doesn't get: code, law, and liberty in cyberspace. Stanf Law Rev 52: 1439–1459
Powell A, Henry N (2017) Sexual violence in a digital age. Palgrave Macmillan, London
Rawls J (1993) Political liberalism. Columbia University Press, New York
Rappaport AJ (2001) Beyond personhood and autonomy: moral theory and the premises of privacy. Utah Law Rev 2001:441–507
Reindenberg JR (2005) Technology and internet jurisdiction. Univ Pa Law Rev 153:1951–1974
Rona G, Aarons L (2016) State responsibility to respect, protect and fulfil human rights obligations in cyberspace. J Natl Secur Law Policy 8:503–530
Ronen Y (2015) Big Brother's little helpers: the right to privacy and the responsibility of internet service providers. Utrecht J Int Eur Law 31:72–86
Roth LM (1999) The right to privacy is political: power, the boundary between public and private, and sexual harassment. Law Soc Inq 24:45–71
Roulin N (2014) The influence of employers' use of social networking websites in selection, online self-promotion, and personality on the likelihood of faux pas postings. Int J Sel Assess 22:80–87
Ross J et al (2019) Sexting coercion as a component of intimate partner Polyvictimization. J Interpers Violence 34:2269–2291
Rowbottom J (2012) To rant, vent and converse: protecting low level digital speech. Camb Law J 71:355–383
Russell D (1998) Pornography, misogony and rape. Sage, London
Rutzick MC (1974) Offensive language and the evolution of first amendment protection. Harv Civ Rights Civ Liberties Law Rev 9:1–28
Saha K et al (2019) Prevalence and psychological effects of hateful speech in online college communities. Proc ACM Web Sci Conf, 2019 June, pp 255–264
Saul J (2006) Pornography, speech acts and context. Proc Aristot Soc 106:229–248
Scanlon T (1979) Freedom of expression and categories of expression. Univ Pittsburgh Law Rev 40:519–550
Schauer F (1993) The phenomenology of speech and harm. Ethics 103:635–653
Schauer F (1998) Internet privacy and the public-private distinction. Jurimetrics 38:555–564
Schwartz PM (1999) Privacy and democracy in cyberspace. Vand Law Rev 52:1609–1701
Scott A et al (2011) Women and the internet: the natural history of a research project. In: Green E, Adam A (eds) Virtual gender: technology, consumption and identity matters. Routledge, London, pp 3–27
Searle J (1969) Speech acts: an essay in the philosophy of language. Cambridge University Press, Cambridge
Sellars AF (2016) Defining hate speech. Berkman Klein Center Research Publication No. 2016-20. Boston University School of Law. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244. Accessed 3 Feb 2022
Siegel A (2020) Online hate speech. In: Persily N, Tucker J (eds) Social media and democracy: the state of the field, prospects for reform. Cambridge University Press, Cambridge, pp 56–88

Sjöholm M (2017) Gender-sensitive norm interpretation by regional human rights law systems. Brill, Leiden
Slane A (2005) Home is where the internet connection is: law, spam and the protection of personal space. Univ Ottawa Law Technol J 2:255–290
Smet S (2010) Freedom of expression and the right to reputation: human rights in conflict. Am Univ Int Law Rev 26:183–236
Solove D (2006) A taxonomy of privacy. Univ Pa Law Rev 154:477–564
Solove D (2007) The future of reputation: gossip, rumor, and privacy on the internet. Yale University Press, New Haven
Solove D (2010) Speech, privacy, and reputation on the internet. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 15–30
Spinello R (2016) Cyberethics: morality and law in cyberspace. Jones & Bartlett Learning, Burlington
Stark C (1997) Is pornography an action? The causal vs. the conceptual view of pornography's harm. Soc Theory Pract 23:277–306
Stone G (2010) Privacy, the first amendment, and the internet. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 174–194
Stone G et al (2020) The first amendment, 6th edn. Aspen Publishing, Frederick
Stoyanova V (2018) Causation between state omission and harm within the framework of positive obligations under the European Convention on Human Rights. Hum Rights Law Rev 18:309–346
Strahilevitz LJ (2005) A social networks theory of privacy. Univ Chic Law Rev 72:919–988
Suzor N et al (2017) Non-consensual porn and the responsibilities of online intermediaries. Melb Univ Law Rev 40:1057–1097
Taddeo M, Floridi L (2016) The debate on the moral responsibility of online service providers. Sci Eng Ethics 22:1575–1603
Taddeo M, Floridi L (2017) New civic responsibilities for online service providers. In: Taddeo M, Floridi L (eds) The responsibilities of online service providers. Springer, Cham, pp 1–10
Taramundi DM (2017) To discriminate in order to fight discrimination paradox or abuse? In: Smet S, Brems E (eds) When human rights clash at the European Court of Human Rights: conflict or harmony? Oxford University Press, Oxford
Tene O, Polonetsky J (2014) A theory of creepy: technology, privacy, and shifting social norms. Yale J Law Technol 16:59–102
Tigwell G, Flatla D (2016) Oh that's what you meant! Reducing emoji misunderstanding. Proceedings of the 18th international conference on human-computer interaction with mobile devices and services adjunct, MobileHCI 2016, pp 859–866. https://www.researchgate.net/publication/307573599_Oh_that's_what_you_meant_reducing_emoji_misunderstanding. Accessed 4 Feb 2022
Tsesis A (2002) Prohibiting incitement on the internet. Va J Law Technol 7:1–41
Tulkens F (2012) When to say is to do: freedom of expression and hate speech in the case-law of the European Court of Human Rights. In: Casadevall J et al (eds) Freedom of expression: essays in honour of Nicolas Bratza. Wolf Legal Publishers, Oisterwijk, pp 279–295
Turner PN (2014) "Harm" and Mill's harm principle. Ethics 124:299–232
Ullmann S, Tomalin M (2020) Quarantining online hate speech: technical and ethical perspectives. Ethics Inf Technol 22:69–80
Velez M (2019) "Why take the photo if you didn't want it online?": agency, transformation, and nonconsensual pornography. Women's Stud Commun 42:452–470
Vitis L, Segrave M (2017) Introduction. In: Segrave M, Vitis L (eds) Gender, technology and violence. Routledge, London
Volokh E (1995) Cheap speech and what it will do. Yale Law J 104:1805–1852

Voorhoof D (2017) Pihl v. Sweden: non-profit blog operator is not liable for defamatory users' comments in case of prompt removal upon notice. https://strasbourgobservers.com/2017/03/20/pihl-v-sweden-non-profit-blog-operator-is-not-liable-for-defamatory-users-comments-in-case-of-prompt-removal-upon-notice/. Accessed 18 Feb 2022

Wajcman J (2004) Technofeminism. Polity Press, Oxford

Waldman AE (2013) Durkheim's internet: social and political theory in online society. N Y Univ J Law Liberty 7:345–430

Weinstein J (1999) Hate speech, pornography, and radical attacks on free speech doctrine. Westview Press, Boulder

Weinstein J (2009) Extreme speech, public order, and democracy: lessons from *The Masses*. In: Hare I, Weinstein J (eds) Extreme speech and democracy. Oxford University Press, Oxford, pp 23–61

Weintraub J (1997) The theory and politics of the public/private distinction. In: Weintraub J, Kumar K (eds) Public and private in thought and practice: perspectives on a grand dichotomy, pp 1–42

White AE (2006) Virtually obscene: the case for an uncensored internet. McFarland Co. Inc., London

Woodzicka JA, Ford TE (2010) A framework for thinking about the (not-so-funny) effects of sexist humor. Eur J Psychol 6:174–195

Yanisky-Ravid S (2014) To read or not to read: privacy within social networks, the entitlement of employees to a virtual private zone, and the balloon theory. Am Univ Law Rev 64:53–104

Chapter 4
Online Gender-Based Offences and International Human Rights Law

4.1 Introduction

As viewed, international human rights law regulations apply equally to the Internet, including both explicit and general provisions encompassing protection against gender-based harm and gender stereotypes. Nevertheless, challenges arise in transposing these standards to the online environment, both in relation to new types of offences and pre-existing violations in online forms. In consideration of this, the following chapter will address whether and how international human rights law applies to select forms of harmful online conduct and speech. This connects to the questions broadly examined in the previous chapter, on the approach to harm, the scope of rights and liability. Feminist legal theories are intersected throughout sections, to consider potential gendered effects arising from gaps, silences and inconsistencies in the regulation of the selected offences in international human rights law *per se*, as well as in the transposition of human rights law provisions to the Internet. In instances of perceived regulatory or interpretative gaps, arguments will be made for a gender-sensitive and functionally equivalent application of human rights law provisions to the Internet.

It should be borne in mind that due to the novelty of ICT-related offences, they are in most cases not defined at the international level and, when transgressions are defined, it is generally outside of the context of the Internet. The headings of the sections are thus in place for the purpose of structuring the analysis. As such, broad categories of offences will be presented, where the specific acts encompassed frequently overlap. For instance, filming and distributing videos of rape or instances of so-called revenge pornography may simultaneously constitute sexual violence, harassment or disclosure of private information. In international human rights law, such offences are violations of various aspects of the right to privacy. I will thus in most instances not attempt to define the offences where definitions do not exist at the international level, but rather discuss the various acts in relation to international human rights law provisions.

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7_4

4.2 Sexual Violence

4.2.1 Introduction

The view on sexuality within international human rights law has moved from being an aspect of the private sphere—beyond the purview of state interference—and a matter of public morals, to being regarded as integral to individual autonomy, requiring international intervention. Sexual relationships engage multiple privacy-related values and it has been affirmed in international human rights law that choices involving sexuality are considered particularly intimate and self-defining.[1] Increasingly, sexuality is explored online, especially by young people who cultivate friendships, flirt and develop their identities. Benefits of the Internet in this respect include the possibility to come to terms with your sexual identity and to find partners with whom to explore sexuality in consensual online/offline fora.[2] Arguably, identity is formed through interactions with other people, institutions and popular culture, as this informs individuals of the various identities at their disposal.[3] The process of understanding your sexual identity generally involves 'fantasizing, experimentation, education, and social interaction' and the Internet has been considered particularly useful in helping LGBTQ individuals understand their sexual orientation.[4] Part of the reason is that the control of family and the community in the development of norms is weakened, allowing for the exploration and development of identity and values at odds with community standards.[5] The most significant aspect of the Internet in this regard is the relative anonymity of users.

At the same time, the Internet is a platform for the invasion of the sexual autonomy of individuals through sexual violence. This includes rape, forced masturbation, forced nudity and image-based sexual abuse. Both the nature and prevalence of sexual violence is gendered, with global statistics indicating that women are disproportionately affected during the course of their lives, both offline and online.[6] While this causes severe individual physical and psychological harm, it also has social effects, by impacting on the ability and incentive of women to act as agents in

[1] *See*, for example, *Dudgeon v the United Kingdom* (1981) 4 EHRR 149, para. 52; *X and Y v the Netherlands* (1986) 8 EHRR 235, para. 27; *Fernandez Ortega* et al. *v Mexico* (preliminary objections, merits, reparations, and costs) IACtHR Series C No 224 (30 August 2010), para. 129. *See* also, generally, Levmore and Nussbaum (2010), p. 10.

[2] Gilden (2016), p. 426.

[3] Long and Chen (2007).

[4] Gilden (2016), pp. 422–423; UN HRC, 'Right to Privacy: Report of the Special Rapporteur on the right to privacy' (16 October 2019) UN Doc. A/HRC/40/63, para. 58.

[5] Gilden (2016), p. 432.

[6] 30% of women have been exposed to sexual violence in a relationship and 7% by a person other than a partner. *See* WHO, 'Global and Regional Estimates of Violence against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-partner Sexual Violence' (2013), p. 2. Victims of image-based sexual abuse are most frequently women and girls. *See* Suzor et al. (2017), p. 1068; Citron and Franks (2014), p. 353.

the public sphere and thus impeding a gender equal Internet. Additionally, the presence of, for example, image-based sexual abuse on the Internet arguably increases the demand for images and videos where individuals are sexually exploited, drawing an analogy to the social harm of child abuse images.[7] Nevertheless, sexual coercion is different online than it is offline through the absence of physical force. Sexual violence through the Internet rarely involves physical interaction between the victim and perpetrator. Conventional approaches to which acts constitute sexual violence and are encompassed within the definitions of particular offences are thus challenged, as is the assessment of harm and the applicability of particular provisions in international human rights law treaties.

4.2.2 The Prohibition and Definition of Sexual Violence

The importance of categorising harm to sexual autonomy as sexual violence is the gravity with which such violations are approached in international human rights law, both in terms of the provisions that apply and the obligations that ensue, requiring extensive state measures of protection. It also affects balancing exercises in conflicts of rights and proportionality assessments. Additionally, the particular harm caused to sexual autonomy is not acknowledged if such offences are categorised as other types of crime. A particular stigma is attached to sexual violence, with the preventive effect plausibly reduced if acts are categorised as, for example, harassment or defamation. It is thus of both a symbolic and legal importance that online harm to sexual autonomy is categorised as sexual violence.

The prohibition on sexual violence is explicitly included in three (regional) human rights law treaties and by way of interpretation in a range of international conventions.[8] Its proscription has also been affirmed by the three regional human rights law courts/commissions, the United Nations Committee on the Elimination of Discrimination against Women (CEDAW Committee), the United Nations Committee Against Torture (CAT), the UN Human Rights Committee (UN HRC) and the United Nations Committee on Economic, Social and Cultural Rights (CESCR), as an aspect of general human rights.[9] It has primarily been considered a violation of the

[7] Citron and Franks (2014), p. 364. This can also be said of adult pornography.

[8] Explicit in Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (the Istanbul Convention) (2011), CETS No. 210, entered into force 1 August 2014 (Art. 36); Inter-American Convention on the Prevention, Punishment and Eradication of Violence Against Women (Belém do Pará Convention) (1994), 33 i.l.m. 1534 (1994), entered into force 3 May 1995 (Art. 2); Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (Maputo Protocol) (2000), Adopted by the second Ordinary Session of the Assembly of the Union, CAB/LEG/66.6, entered into force 25 November 2005 (Art. 3 (4), Art 11 (3), Art, 22 (b), Art. 23 (b)).

[9] UNCESCR, 'General Comment No. 22 on the Right to Sexual and Reproductive Health' (2016), UN Doc. E/C.12/GC/22, para. 29. *See* the following fn for the other sources.

prohibition on torture, inhuman or degrading treatment and the right to privacy.[10] It is additionally considered a form of gender discrimination. Regional treaties on gender-based violence, soft law documents—including general recommendations and views issued by UN treaty bodies—and case law from certain regional human rights law courts recognise that sexual violence is a gendered offence, both quantitatively and qualitatively.[11] This pertains to the act of sexual violence, the social norms underlying the offence or state negligence in preventing the crime. The fact that the *form* of violation perpetrated against women in a multitude of settings often involves sexual assault and sexual harassment is seen as a gendering aspect of the crime.[12] For example, the IACmR has affirmed that rape has '. . .specific gender-specific causes and consequences [and it is] used to submit and humiliate and as a method of destroying the autonomy of the woman'.[13]

This comprehensive denunciation demonstrates the broad impact of sexual violence on individuals and society at large, implicating both civil and political rights as well as economic, social and cultural rights. At a general level, the prohibition on sexual violence remains the same *vis-à-vis* the Internet. This is evident not only through the general assertion of human rights law applying equally online, but also more specifically involving sexual violence on the Internet. [14]

[10] *M. C. v Bulgaria* (2005) 40 EHRR 20; *Fernandez Ortega* et al. *v Mexico* (IACtHR); *Equality Now and Ethiopian Women Lawyers Association (EWLA) v Federal Republic of Ethiopia,* ACmHPR, Communication No. 341/2007, Adopted by the African Commission on Human and Peoples' Rights during the 19th Extra-Ordinary Session, from 16 to 25 February 2016, Banjul, The Gambia; *Fulmati Nyaya v Nepal* Communication No. 2556/2015, UN HRC, UN Doc. CCPR/C/125/D/2556/2015 (11 June 2019); *Mrs. A v Bosnia and Herzegovina*, Communication No. 854/2017, UNCAT, UN Doc. CAT/C/67/D/854/2017 (22 August 2019).

[11] *See* UNGA, 'Declaration on the Elimination of Violence against Women' (23 February 1994) UN Doc. A/RES/48/104, (Art. 2); the Maputo Protocol (Art. 1 (j)); the Istanbul Convention (Art. 3 (a)); the Belém do Pará Convention (Art. 2); CEDAW, 'General Recommendation No. 19 on Violence Against Women' (1992) UN Doc A/47/38, para. 6; UNCHR, 'Report of the Special Rapporteur on Violence against Women, its Causes and Consequences, Yakin Ertürk' (26 December 2003) UN Doc. E/CN.4/2004/66, para. 35; *Fernandez Ortega* et al. *v Mexico* (IACtHR); *Case of Miguel Castro-Castro Prison v Peru* (merits, reparations and costs) IACtHR Series C No. 160 (25 November 2006); *Case Egyptian Initiative for Personal Rights and Interights v Egypt*, ACmHPR, Communication No. 323/06 (1 March 2011). Forced nudity has similarly been categorised as a form of sex discrimination in certain instances. *See Abramova v Belarus*, CEDAW Communication No. 23/2009, UN Doc. CEDAW/C/49/D/2009 (27 September 2011), para. 7.2.

[12] UNCHR, 'Report of the Special Rapporteur on Torture and Cruel, Inhuman or Degrading Treatment or Punishment, Mr. Nigel S. Rodley, submitted pursuant to Commission on Human Rights resolution 1992/32' (12 January 1995) UN Doc. E/CN.4/1995/34, para. 16.

[13] Relayed in *Rosendo Cantú* et al. *v Mexico* (preliminary objections, merits, reparations, and costs) IACtHR Series C No. 216 (31 August 2010), para. 81.

[14] *K.U. v Finland* (2009) 48 EHRR 52; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19' (14 July 2017) UN Doc. CEDAW/C/GC/35, paras. 14 and 20; UNHRC, '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/38/4, para. 27.

However, most international human rights law treaties are silent on the scope of the concept of sexual violence and it has not been defined in the case law of international human rights law bodies. In contrast, specific acts of sexual violence have on an *ad hoc* basis been delineated, for example, by regional human rights law courts, mainly involving the elements of the crime of rape.[15] Nevertheless, it is clear that sexual violence encompasses a range of acts that violate individual sexual autonomy. For example, state obligations provided in the CoE Convention on Preventing and Combating Violence against Women and Domestic Violence (Istanbul Convention) *vis-à-vis* sexual violence indicate that it involves (a) engaging in non-consensual vaginal, anal or oral penetration of a sexual nature of the body of another person with any bodily part or object; (b) engaging in other non-consensual acts of a sexual nature with a person; and (c) causing another person to engage in non-consensual acts of a sexual nature with a third person.[16] Although this includes various physical acts, it does not extend to speech, and sexual harassment is regulated in a separate provision of the Convention.[17] Similarly, although sexual harassment, including physical acts, is considered a form of gender-based violence in regional human rights law treaties and by the CEDAW Committee, it is generally not considered synonymous with sexual violence.[18]

This does not entail that physical contact is a requirement. Under specific circumstances, forced nudity constitutes sexual violence and has been addressed as such in both international human rights law and international criminal law. [19] For example, in categorising forced nudity as sexual violence, the Inter-American Court of Human Rights (IACtHR) relied on the broad definition of sexual violence established by the International Criminal Tribunal for Rwanda (ICTR) in the *Akayesu* case, encompassing '. . .actions with a sexual nature committed with a

[15] *M. C. v Bulgaria* (ECtHR); *Miguel Castro-Castro Prison v Peru* (IACtHR).

[16] Art. 36 of the Istanbul Convention.

[17] Art. 40 of the Istanbul Convention.

[18] Art. 2 of the Belém do Pará Convention; Arts. 12(1) (c) and 13(c) (d) of the Maputo Protocol; Art. 40 of the Istanbul Convention; CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 17; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 14.

[19] *See,* for example, *Prosecutor v Jean-Paul Akayesu*, ICTR, Case No. ICTR-96-4-T (2 September 1998), para. 697; *Abramova v Belarus* (CEDAW); *Miguel Castro-Castro Prison v Peru* (IACtHR). The UN Special Rapporteur on Torture considers 'being stripped naked, invasive body searches' a form of sexual violence. *See* UNHRC, 'Report of the Special Rapporteur on Torture and other Cruel, Inhuman or Degrading Treatment or Punishment, Manfred Nowak' (15 January 2008) UN Doc. A/HRC/7/3, para. 34; CEDAW, 'Concluding Observations on India' (22 October 2010) UN Doc. CEDAW/C/IND/CO/SP.1, para. 5. It is also criminalised in the Rome Statute of the ICC, as a form of sexual violence under the chapeau of crimes against humanity (Art. 7 (1) (g)-6) and war crimes (Art. 8 (2) (b) (xxii)-6) of the Rome Statute of the International Criminal Court (1998) 2187 UNTS 90, entered into force 1 July 2002. For a discussion on this issue in relation to the Rome Statute, *see* 'ICC Policy Paper on Sexual and Gender-Based Crimes', The Office of the Prosecutor, ICC (June 2014) <https://www.icc-cpi.int/iccdocs/otp/otp-Policy-Paper-on-Sexual-and-Gender-Based-Crimes%2D%2DJune-2014.pdf> Accessed 14 February 2022, para. 17.

person without their consent, which besides including the physical invasion of the human body, may include acts that do not imply penetration or even any physical contact whatsoever'.[20] It should be noted that while threats of rape have been considered violations of international human rights law, they have not been construed as sexual violence *per se*, further discussed in Sect. 4.3.3. Meanwhile, multiple definitions of sexual violence have been developed in soft law documents. Although the definitions contain some variations, it is similarly clear that sexual violence involves a spectrum of acts, both physical and non-physical, where certain acts such as rape are considered graver than others. In certain instances, the definitions encompass speech, such as unwanted sexual comments.[21]

Sexual violence has to a greater extent been defined by courts and tribunals in the context of international criminal law. This also encompasses a range of acts of a sexual nature, involving parts of the body commonly associated with sexuality, which infringe on the physical and psychological integrity of the victim.[22] In case law, rape, forced nudity, sexual molestation and sexual slavery have been categorised as forms of sexual violence, that is, both physical and non-physical acts with a sexual element, although not exclusively speech-based offences.[23] Although the context of ICL has a certain bearing on the definition and understanding of what constitutes sexual violence, the contextual differences should not have an impact on the determination of what type of acts violate individual sexual autonomy.[24]

In terms of *online* sexual violence, as noted previously, acts involving children have been specifically addressed at the international level, encompassed by provisions on sexual exploitation in the United Nations Convention on the Rights of the

[20] *Miguel Castro-Castro Prison v Peru* (IACtHR), para. 306.

[21] Sexual violence is defined by WHO as 'any sexual act, attempt to obtain a sexual act, unwanted sexual comments or advances, or acts to traffic. . .against a person's sexuality using coercion. . .'. *See* WHO, Krug et al. (eds), 'World Report on Violence and Health' (2002) <http://apps.who.int/iris/bitstream/handle/10665/42495/9241545615_eng.pdf;jsessionid=9E9BAD229FBECC80E4E32AB0308AF786?sequence=1> Accessed 30 March 2022, p. 149. The ACmPHR defines sexual violence as including sexual harassment, forced nudity, forced pornography and forced masturbation and any other forced touching that the victim is compelled to perform on himself/herself or a third person. *See* ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (adopted by the African Commission on Human and Peoples' Rights during its 60th Ordinary Session held in Niamey, Niger from 8 to 22 May 2017), paras. 3 (1) (a) and (b). *See* also UN Commission on Human Rights, 'Final Report Submitted by Ms. Gay J. McDougall, Special Rapporteur on the Situation of Systematic Rape, Sexual Slavery and Slavery-Like Practices during Armed Conflicts' (22 June 1998) UN Doc. E/CN.4/Sub.2/1998/13, paras. 21–22.

[22] *Prosecutor v Jean-Paul Akayesu*, ICTR, Case No. ICTR-96-4-T (2 September 1998), para. 688.

[23] ibid. (rape, forced nudity); *Prosecutor v Miroslav Kvocka*, ICTY, Case No. IT-98-30/1-T, Judgment of 2 November 2001, para. 180 (includes sexual slavery, molestation). *See* also the Rome Statute: Art. 7 (1) (g), Art. 8 (2) (b) (xxii), Art. 8 (2) (e) (vi).

[24] Not only does this pertain to the legal requirements of, for example, a nexus to an armed conflict in relation to war crimes, but it may also have an impact on the nature and purpose of the crime itself, and the interpretation of the elements of the crime, such as non-consent and coercion.

Child (CRC), [25] the CoE Convention on Protection of Children against Sexual Exploitation and Sexual Abuse (Lanzarote Convention),[26] in addition to several other treaties[27] and EU-regulations.[28] This includes the prohibition on sexual exploitation through prostitution, performing live sexual acts through streaming services and the production of child sexual abuse material, for example, by recording online abuse.

Meanwhile, whereas certain forms of online violence disproportionately affecting women have been categorised as being gender-based in international human rights law, there is no comprehensive delineation of which acts committed on the Internet against adults constitute sexual violence. In the main, the applicability of existing concepts to the online sphere has been affirmed, for example, the definition in the Istanbul Convention.[29] This clearly extends to instances of physical sexual violence perpetrated through the Internet. Nevertheless, although the approach in international human rights law is that sexual violence cannot be limited to certain physical acts—encompassing a range of harmful constraints on a person's sexual autonomy—the application of general provisions to the Internet risks excluding certain forms of online sexual violence since speech-based offences, such as verbal sexual harassment, seldom are encompassed in the concept of sexual violence.

Certain international bodies have on an *ad hoc* basis affirmed the categorisation of sextortion and image-based sexual abuse as sexual violence.[30] The latter includes

[25] Art. 34 (c) of the United Nations Convention on the Rights of the Child (1989) GA Res 44/25, annex, 44 UN GAOR Supp (No 49) at 167, UN Doc. A/44/49 (1989), 1577 UNTS 3, entered into force 2 September 1990 (sexual exploitation); The Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography, GA Res. 54/263, Annex II, 54 UN GAOR Supp (No. 49) at 6, UN Doc. A/54/49, Vol. III (2000), entered into force January 18, 2002; CRC Committee, 'Guidelines regarding the implementation of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography' (10 September 2019) UN Doc. CRC/C/156, para. 41.

[26] Art. 18 (sexual exploitation), Art. 20–21 (child pornography), Art. 23 (grooming) of the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (Lanzarote Convention) (ETS No. 201) 25 October 2007.

[27] Art. 27 (c) of the African Charter on the Rights and Welfare of the Child (ACRWC) OAU Doc. CAB/LEG/24.9/49 (1990), entered into force Nov. 29, 1999; Art. 3 (b) of ILO C182—Worst Forms of Child Labour Convention, 1999 (No. 182).

[28] Art. 2 (e) of Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA (2011) OJ L335/1.

[29] CoE, Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', adopted on 20 October 2021; European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', COM (2022) 105 final (8 March 2022), p. 3.

[30] European Parliament, Directorate-General for Internal Policies of the Union, Wilk, A., 'Cyber violence and hate speech online against women' (2018), p. 17; OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', Prepared by the General Secretariat of the OAS (2021) OEA/Ser.D/XXV.25, p. 42.

(1) the *creation* of nude or sexual images without consent (including hidden recordings of consensual sex, nudity or sexual assault; doctored photographs with a superimposed face on pornographic images; and hacked images and photographs taken when the victim was asleep/affected by drugs or alcohol); (2) the *distribution* without consent (in cases where the photographs were taken with or without consent); or (3) the *threat* of distribution of images.[31] Image-based sexual abuse is in this manner distinguished from other forms of unlawful disclosures of private information, by involving sexual content. Although the purpose may vary—including relationship retribution, sextortion, blackmail, voyeurism, sexual gratification, financial gain or enhancing one's social status -[32] the sexuality of an individual is exposed and controlled and it limits the victim's ability to autonomously make decisions on matters relating to his/her sexuality.[33] Thus, as the acts are sexual and non-consensual in nature, they transgress the protective values of the prohibition on sexual violence.

In particular so-called "revenge porn" is increasingly recognised as a form of sexual violence, involving the distribution of intimate images without consent, although it also falls within the broader scope of cyber harassment and gendered hate speech.[34] For example, the Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO) of the CoE has considered this form of image-based sexual abuse as encompassed within the definition of sexual harassment in the Istanbul Convention, rather than sexual violence, while noting that it also overlaps with sexist hate speech.[35] In certain instances, sextortion and image-based sexual abuse are categorised as violence against women, but not as a specific type of

[31] Although several concepts may be employed to comprehensively acknowledge such acts as sexual violence, the term "image-based sexual abuse" will be employed in the following, as proposed by scholars Clare McGlynn and Erika Rackley. *See* McGlynn and Rackley (2017). *See* also Powell and Henry (2017), p. 118 and the definition in Art. 7 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[32] Powell and Henry (2017), pp. 121–130; UNHRC, '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), para. 41. It should in this regard be noted that the term "revenge porn" is rejected by several scholars, as it does not address circumstances outside of intimate relationships and may encourage victim blaming through the presumption of wrongdoing on the part of the victim. Additionally, by categorising it as a form of pornography, the harm to the victim is minimised, by likening the images to an acceptable genre of publication. This is a similar argument as in relation to "child pornography", which is increasingly addressed as "child exploitation material". *See* Powell and Henry (2016), p. 400. However, children are exploited through photos being taken, which is not necessarily the case in revenge porn, where the capturing of the image may be consensual.

[33] Roth (1999), p. 57; Citron and Franks (2014), p. 353.

[34] UNHRC, '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), para. 41. *See* also Powell and Henry (2017), p. 119; Suzor et al. (2017), p. 1092.

[35] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', paras. 38–39.

violation.[36] For example, the proposal for an EU directive on violence against women aims to harmonise the criminalisation of "non-consensual sharing of intimate or manipulated material"—categorised as cyber violence—albeit not specifically considered sexual violence.[37]

The categorisation of image-based abuse as sexual violence is also contested. Arguably, it is unsupported by domestic or international law, given its lack of physical interaction, despite the inclusion of, for example, forced nudity in the concept.[38] From this viewpoint, when such acts are part of, for example, sextortion, they may reach the threshold of sexual violence. This concerns threats of disclosure of intimate images in particular. Similarly, in instances of acts encompassed within the scope of image-based sexual abuse the approach by, for example, the ECtHR has been inconsistent. Albeit the ECtHR in *Söderman v Sweden* and *Khadija Ismayilova v Azerbaijan* considered the illicit recording of a child in the nude and the covert taking of photographs and filming of sexual acts, respectively, as violations of sexual autonomy, these acts were not explicitly categorised as sexual violence.[39]

Broader approaches to online sexual violence have mainly been advanced by scholars, including the concept of technology-facilitated sexual violence.[40] This comprises both physical and virtual acts facilitated by information and communication technologies (ICTs), including the sending of nude photographs (so-called dick pics) as a form of indecent exposure.[41] However, the latter offence is most commonly categorised as sexual harassment, which appears to be linked to the gravity of harm, and will be addressed in Sect. 4.3.2. These non-physical intrusions on sexual autonomy online thus highlight the unclear boundaries between sexual violence and sexual harassment in international human rights law.

4.2.3 Hierarchies of Sexual Violence

Although sexual violence involves a range of acts, these are not considered equally grave. Case law from the regional human rights law courts provide further insight into the hierarchies of different types of sexual violence, applicable norms and the content of state obligations. More specifically, the ECtHR considers that sexual

[36] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 42; UNHRC, '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), paras. 34–35.

[37] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', Art. 7.

[38] Patton (2015), p. 429.

[39] *Söderman v Sweden* (2014) 58 EHRR 36; *Khadija Ismayilova v Azerbaijan* App no 30778/15 (ECtHR, 27 February 2020).

[40] Powell and Henry (2017).

[41] Powell and Henry (2017), p. 120.

violence, with acts ranging from sexual molestation to rape, falls within the scope of Article 3 and Article 8 of the Convention, depending on the gravity of the act. That is, it involves violations of either the prohibition on torture, inhuman or degrading treatment or the right to privacy. Meanwhile, sexual violence is seen as an inherently debasing violation and an affront to essential aspects of private life.[42] The harm is thus recognised as an objective factor.

Several hierarchies are noticeable in the case law on sexual violence, one being the Court's distinction between physical acts of sexual violence—constituting violations of Article 3—and non-physical harm—falling within the scope of Article 8. In *Söderman v Sweden*, the Court held that since no physical harm had been suffered by the victim through being filming naked, it did not reach the requisite severity level of Article 3, despite the fact that the Court found the circumstances particularly aggravating with the offence occurring in the home of the victim by a person of trust.[43] Similarly, the placement of an advertisement of a child on a dating website for sexual purposes led the Court in *K.U. v Finland* to conclude that the state in question had failed to protect the child's right to privacy by not ensuring means of identifying the perpetrator.[44] Furthermore, although forced nudity under limited circumstances—involving state actors and prolonged periods of having to remain nude—has been considered a violation of the prohibition on inhuman or degrading treatment, it has also in the main been categorised as an intrusion on the sexual autonomy of a person and thus a violation of the right to privacy.[45] It appears that the domestic qualification of non-consensual sexual acts is not of relevance to the ECtHR, that is, whether categorised as rape, sexual assault or molestation, as long as the constitutive acts are prohibited in domestic criminal law, providing effective protection.[46] However, the gravity of the act has an impact on the categorisation of the violation and the content of obligations.

In the jurisprudence of the ECtHR, a distinction is furthermore made with regard to the identity of the assailant, with violations of the prohibition on torture requiring the perpetration by a state actor.[47] In early case law, rape committed by private

[42] *X and Y v the Netherlands* (1986) 8 EHRR 235, para. 27.

[43] *Söderman v Sweden* (ECtHR).

[44] *K.U. v Finland* (ECtHR).

[45] *Miguel Castro-Castro Prison v Peru* (IACtHR); *Abramova v Belarus* (CEDAW); UN HRC, 'Report of the Special Rapporteur on Torture and Cruel, Inhuman or Degrading Treatment or Punishment, Juan E. Méndez, Mission to Mexico' (29 December 2014) UN Doc. A/HRC/28/68/Add.3, para. 28; ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)' (Adopted at the 21st Extra-Ordinary Session of the African Commission on Human and Peoples' Rights, held from 23 February to 4 March 2017 in Banjul, The Gambia), para. 58; ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), paras. 3 (1) (a) and (b).

[46] *C. A. S. and C. S. v Romania* (2015) 61 EHRR 18.

[47] The Court in its earliest case on rape committed by a state actor considered it to constitute inhuman treatment, while in current case law it is primarily categorised as a form of torture. *See*

individuals was categorised as violations of Article 8.[48] More recently, a clear development is noticeable in the Court's jurisprudence in viewing the harm of sexual violence with increased severity, more commonly categorising rape by non-state actors as violations of Article 3 rather than Article 8, although still requiring a state perpetrator in relation to torture. Meanwhile, in the *M.C. v Bulgaria* case, the ECtHR considered that the alleged rape violated the applicant's right to autonomy under Article 8 as well as physical and mental integrity under Article 3.[49] Different aspects of sexual violence may accordingly be acknowledged through the application of the different articles. The distinction between sexual violence as a violation of Article 3 as opposed to Article 8 thus appears to be both an evolutive development as well as a substantive one. Although there is no formal hierarchy of rights, Article 3 provides stricter state obligations in that the provision does not allow for limitations or derogations unlike Article 8, which is a qualified right. As noted in *Y. v Slovenia*, charges of violations of Article 3, as opposed to Article 8, require 'particularly thorough scrutiny' by the Court.[50] It is thus of both a symbolic and practical significance.

Although not involving as extensive case law on the topic, an analogous delineation is evident in other regional human rights law systems[51] as well as by UN treaty bodies. [52] The case law predominantly involves rape, with a few cases

---

Cyprus v Turkey (1982) 4 EHRR 482 (inhuman or degrading treatment); *Maslova and Nalbandov v Russia* (2009) 48 EHRR 37 (torture); *Aydin v Turkey* (1998) 25 EHRR 251 (torture).

[48] *X and Y v the Netherlands* (ECtHR).

[49] *M. C. v Bulgaria* (ECtHR). This was apparent in *Y. v Slovenia* (2016) 62 EHRR, where various parts of the domestic court proceedings in a case of rape were analysed under separate articles: impartiality and delay in proceedings under Art. 3 and humiliation associated with the proceedings under Art. 8.

[50] *Y. v Slovenia* (2016) 62 EHRR 3.

[51] Most cases in the Inter-American human rights system involve sexual assault committed by state actors, categorised as torture and a violation of the right to privacy. *See*, for example, *Raquel Martí de Mejía v Perú*, IACmHR, Case 10.970, Report No. 5/96, OEA/Ser.L/V/II.91 Doc. 7 at 157 (1 March 1996); *Rosendo Cantú* et al. *v Mexico* (IACtHR). Solely a few cases involve sexual violence perpetrated by a private individual. These have primarily been addressed as violations of the right to private life as well as equality and non-discrimination. *See V. R. P. and V. P. C. v Nicaragua* (preliminary objections, merits, reparations and costs) IACtHR Series C No. 350 (8 March 2018). ACmHPR case law on sexual violence is limited but includes the act of rape, forced marriage and sexual harassment. *See*, for example, *Equality Now and Ethiopian Women Laywers Association (EWLA) v Federal Republic of Ethiopia,* ACmHPR, Communication No. 341/2007, Adopted by the African Commission on Human and Peoples' Rights during the 19th Extra-Ordinary Session, from 16 to 25 February 2016, Banjul, The Gambia. For a general overview, *see* Sjöholm (2017).

[52] The CAT, the CEDAW Committee, the UN HRC and the UN Special Rapporteur on Torture indicate that rape under certain circumstances may amount to torture when involving a state actor and, in instances involving non-state actors, inhuman or degrading treatment, as well as violations of the right to privacy and non-discrimination. *See Vertido v the Philippines*, CEDAW Communication No. 18/2008, UN Doc. CEDAW/C/46/D/18/2008 (16 July 2010), para. 8.7 (discrimination); *L.N.P. v Argentina*, Communication No. 1610/2007, UNHRC, UN Doc. CCPR/C/102/D/1610/2007 (16 August 2011), paras. 13.3 and 13.7 (privacy, discrimination); *Fulmati Nyaya v Nepal*

concerning forced nudity and sexual harassment, affirming a similar hierarchy in terms of gravity and applicable provisions. Nevertheless, the IACtHR in *López Soto and Others v Venezuela* categorised sexual slavery, including rape, perpetrated by a non-state actor as a form of torture.[53]

Given the novelty of image-based sexual abuse, it has not been addressed in relation to specific rights in international human rights law. However, it is clear that the right to privacy is applicable, with image-based sexual abuse engaging two aspects of this right. Non-consensual recordings and/or the distribution of intimate sexual content transgress principles of secrecy and cause reputational harm. In addition, as such abuse violates the victim's sexual autonomy, it is also a form of sexual violence, regardless of the lack of physical contact.[54] It thus comprises an intrusion on both the individual and social dimensions of the right to privacy.

The assessment of the *recording* of nude or intimate images in relation to the right to privacy centres on the consent of the person involved. As noted previously, the ECtHR has affirmed that capturing an image of a person without consent may involve an invasion of privacy, depending on whether it occurs in the private or public sphere, and the reasonable expectations of privacy *vis-à-vis* the latter.[55] This protection arises regardless of the content of the photograph. However, where the picture or recording contains intimate or sexual content, it is considered particularly invasive and also transgresses sexual autonomy, as noted in *Söderman v Sweden*.[56] Furthermore, in *Khadija Ismayilova v Azerbaijan* of the ECtHR, the capturing of still images and videos by a hidden camera in the bedroom of a journalist engaging in sexual intercourse with her boyfriend, were considered 'an affront to human dignity', involving a 'serious, flagrant and extraordinarily intense invasion of her private life'.[57]

In instances of so-called revenge pornography, the victim may have consented to an image being taken but not to the subsequent dissemination, which is still within the realm of the offence.[58] The underlying act is thus not necessarily illegal. In such cases, the *distribution* alone constitutes an offence. The case law discussed in the section on disclosure of private information is applicable, although a distinction can

(UN HRC), para. 8 (torture/inhuman or degrading treatment, privacy, discrimination); *Mrs. A v Bosnia and Herzegovina* (UNCAT), para. 7.3 (torture). Nudity: *Abramova v Belarus*, CEDAW, para. 7.2; *Miguel Castro-Castro Prison v Peru* (merits, reparations and costs) IACtHR. Harassment: *Egyptian Initiative for Personal Rights and Interights v Egypt* (ACmHPR), para. 165, 202.

[53] *López Soto and others v Venezuela* (merits, reparations and costs) IACtHR Series C No 362(14 May 2019), para. 188.

[54] Citron and Franks (2014), p. 362.

[55] *Reklos and Davourlis v Greece* App no 1234/05 (ECtHR, 15 January 2009), para. 40; *Peck v United Kingdom* (2003) 36 EHRR 41, para. 59.

[56] *Söderman v Sweden* (ECtHR), paras. 51–52, 82, 86.

[57] *Khadija Ismayilova v Azerbaijan* (ECtHR), para. 116.

[58] Powell and Henry (2017), p. 120. In their survey of Australia and the UK, approximately 9% of women in Australia had a nude image taken without consent and 10% in the UK. 8% resp. 9.7% were threatened to post such.

be made, given the additional intrusion on sexual autonomy. In the cases of *K.U. v Finland* and *Volodina v Russia*, the negligence by the state *vis-à-vis* the publication of private information, in the former case involving contact information published on a dating website, and in the latter, private images disseminated on social media, were in violation of the European Convention on the Protection of Human Rights and Fundamental Freedoms (ECHR).[59] Whereas the former involved a violation of the right to privacy, the dissemination in the latter case was addressed in conjunction with other acts of domestic violence and thus contravened the prohibition on inhuman treatment. In *Khadija Ismayilova v Azerbaijan*, the intimate videos were posted online. While the identity of the publishers remained unknown, positive obligations to protect her privacy arose also in this regard.[60]

It should be noted that a specific category of recording and/or distributing intimate images is content capturing instances of sexual assault. It includes the distribution of both pre-recorded images and the live streaming of assault, for example, through social media. Victims of sexual assault may experience heightened levels of trauma and humiliation, further exacerbating the emotional harm of sexual violence. It is considered an expression of masculine entitlement in the form of control and additional abuse, which may affect the categorisation.[61] For example, it is listed as an aggravating circumstance in instances of rape in the proposed EU directive on violence against women.[62] Meanwhile, *threats* to distribute intimate images will be addressed in Sect. 4.3.3.

It is thus apparent that although all these acts of sexual violence are prohibited, a hierarchy based on the gravity is generally upheld, affecting the applicability of provisions. Physical acts of sexual violence, especially involving penetration, are considered more severe than non-physical acts, such as filming a person in the nude. Sexual violence perpetrated by a state actor is in turn considered graver than that performed by a non-state actor. As noted above, a gradation of harm is embedded in international human rights law, in view of the hierarchy of rights and differences in state obligations. Recognising varying levels of severity attached to different types of sexual violence is thus not controversial *per se*. Nevertheless, the current approach to the gradation of sexual violence in international human rights law has particularly detrimental effects in relation to online acts. Considering the aim of effectively protecting individual sexual autonomy and ensuring that rights apply equally online and offline, a re-examination of current standards in light of the online environment is thus necessary.

[59] *K.U. v Finland* (ECtHR); *Volodina v Russia* App no 31261/17 (ECtHR, 9 July 2019).

[60] *Khadija Ismayilova v Azerbaijan* (ECtHR), para. 78.

[61] Powell and Henry (2017), p. 118.

[62] Art. 13 (n) European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

4.2.4 Obligations to Protect

4.2.4.1 Criminalisation

Positive obligations have been developed by international human rights law bodies on measures to protect individuals against sexual violence, including legislative efforts. A certain distinction is made depending on the type of sexual violence. The ECtHR and UN treaty bodies clearly require that states adopt *criminal* laws prohibiting rape, rather than civil law legislation, as the latter is not sufficient from the perspective of effectively deterring such grave crimes.[63] This has similarly been affirmed in relation to other forms of sexual violence, including forced nudity and recording and publishing nude images online without the consent of the person.[64] In contrast, either criminal or civil law measures were considered appropriate in the case of filming a person in the nude, given the lack of physical violence.[65] Meanwhile, according to the CEDAW Committee and the UN Special Rapporteur on Violence against Women, criminalisation is required in relation to revenge pornography.[66] It is unclear whether this obligation extends to other acts categorised as image-based sexual abuse, such as doctored images and threats of distribution, and the non-consensual receipt of nude images. However, the proposal for an EU directive, including cyber violence against women, is a first step towards harmonising the criminalisation of image-based sexual abuse, including the production and distribution of intimate images, as well as threats of such acts.[67]

The obligation to criminalise involves a qualitative element, in that such laws must be *effective*. Although states maintain a margin of appreciation in defining offences at the domestic level, international human rights law bodies increasingly direct states in the required elements of crimes, especially involving rape.[68] Narrow

[63] As seen in *X and Y v the Netherlands* (ECtHR), para. 27; *M. C. v Bulgaria* (ECtHR), para. 186; Art. 36 of the Istanbul Convention; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 33; UN HRC, 'Rape as a grave, systematic and widespread human rights violation, a crime and a manifestation of gender-based violence against women and girls, and its prevention: Report of the Special Rapporteur on violence against women, its causes and consequences, Dubravka Šimonović' (19 April 2021) UN Doc. A/HRC/47/26.

[64] *Khadija Ismayilova v Azerbaijan* (ECtHR); *Miguel Castro-Castro Prison v Peru* (IACtHR) (as an aspect of inhuman or degrading treatment).

[65] *Söderman v Sweden* (ECtHR).

[66] The UN Special Rapporteur on Violence against Women has categorised "revenge porn" as a form of violence against women and called for criminalisation as a means of protection. *See* UNHRC, '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), paras. 33, 101. The CEDAW Committee has also specifically commended states for criminalising online revenge porn. *See*, for example, CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017) UN Doc. CEDAW/C/NOR/CO/9, para. 22.

[67] Art. 7, European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[68] *See*, for example, *M. C. v Bulgaria* (ECtHR).

definitions and interpretations of the elements of specific offences, particularly when affirming gender stereotypes, have been considered discriminatory as they often affect female victims more severely.[69] An issue arising in the context of technology-facilitated sexual violence is thus whether established international human rights law standards requiring certain definitions or criminal elements translate to the Internet.

The scope of obligations to criminalise will in the following section focus on the definition of rape, forced masturbation and forced nudity, given the fuller development *vis-à-vis* such forms of sexual violence in international human rights law. Nevertheless, it can be noted at a general level in relation to image-based sexual abuse that few states have explicitly criminalised such acts, although regulation of the non-consensual distribution of intimate images in particular is increasing.[70] More commonly, states apply existing provisions on defamation, the public disclosure of private information, harassment, stalking, copyright laws or the dissemination of child pornography.[71] For example, in a domestic case in Sweden involving a gang rape live streamed on Facebook and published on Snapchat, the charge for streaming the video was aggravated defamation.[72] Such regulation thus often involves civil law, is often narrow in scope—excluding most instances of image-based sexual abuse—and can thus be viewed as ineffective in preventing such offences.[73]

4.2.4.1.1 Delineating Rape, Forced Masturbation and Forced Nudity

The prohibition of rape has been affirmed as constituting customary international law in International Humanitarian Law (IHL), ICL and international human rights law.[74] Despite this comprehensive reproach and the near universal criminalisation of rape in domestic law, challenges arise in defining the offence and transposing

[69] The CEDAW Committee has criticised the substance of domestic laws in relation to rape. For example, *R.P.B. v. The Philippines*, CEDAW Communication No. 34/2011, UN Doc. CEDAW/C/57/D/34/2011 (12 March 2014); *V.P.P. v. Bulgaria* (CEDAW); *Vertido v the Philippines* (CEDAW).

[70] UNHRC, '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), para. 82.

[71] ibid., para. 81. Several states have adopted laws on revenge porn, for example, involving the disclosure of private sexual photographs or films without the consent of the person depicted, with the intention of causing the victim distress. Domestic cases include *Police v Ravshan Usmanov [2011] NSWLC 40* of Australia, in which a man was convicted of posting naked photographs of an ex-girlfriend on Facebook.

[72] NJA 2018 s. 562, Judgment on 2018-07-02, Högsta Domstolen.

[73] For example, copyright laws tend to apply solely to instances where the victim is the author of the image. Criminal laws on stalking concern repeated behaviour. When specifically criminalised, definitions often require intent to cause harm—which may be difficult to prove—or exclude threats to release an image, which limits the possibility to prevent such forms of abuse.

[74] Henckaerts and Doswald-Beck, for the ICRC (2005), p. 323. This applies also to ICL. *See* also CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para.

elements of the crime to the context of the Internet.[75] Whether sexual coercion through ICTs may be categorised as rape has not been considered at the international level. Thus, the following section will provide an overview of the international approach to defining rape beyond this specific context and subsequently consider its possible application to the digital sphere.

The definition of rape consists of several elements. The focus is commonly on non-consent, coercion or the use of force at both the domestic and the international level. The elements of the crime correlate with the perceived harm, broadly understood to be either the illegitimate use of force resulting in physical injury or the constraint of the sexual autonomy of the victim, perpetrated through a broad range of non-consensual forms of pressure that do not require the use of physical force.[76] Different forms of psychological pressure may be included in a non-consent based standard, such as threats of violence or withholding benefits. However, physical harm and restrictions on autonomy are not mutually exclusive, as such a dichotomy implies. Consent is an abstract concept that may be understood as either attitudinal or performative, that is, involving the internal preferences of an individual or the requirement of physical displays of such preferences.[77] It should in this regard be noted that feminist scholars to an extent disagree as to the harm of rape and, connected to this aspect, the elements of non-consent and force.[78] Largely, a non-consent based standard is favored, since force is not considered to be reflective of harm to a person's autonomy. This element would also more adequately recognise online sexual violence.

In international human rights law, there is a general consensus in prohibiting non-consensual sexual acts, that is, a definition centred on the non-consent of the victim. The ECtHR, the IACtHR and the CEDAW Committee have all affirmed such an approach in case law. It is also explicitly included in the Istanbul Convention.[79] According to the ECtHR in *M.C. v Bulgaria*, in view of sources in both international human rights law and international criminal law, the definition of rape and other forms of sexual violence must in practice focus on non-consent, rather than force or the threat of force, in order to effectively protect individuals against sexual

2 (which includes sexual violence). Rape can also be a form of torture or genocide, which are prohibited in customary international law.

[75] UN HRC, 'Rape as a grave, systematic and widespread human rights violation, a crime and a manifestation of gender-based violence against women and girls, and its prevention: Report of the Special Rapporteur on violence against women, its causes and consequences, Dubravka Šimonović' (19 April 2021).

[76] Eriksson (2011), pp. 58–59.

[77] It may encompass not just verbal agreement but also be construed as the absence of refusal or resistance. *See* Westen (2016), ch. 1.

[78] For an overview, *see* Brenner (2013).

[79] Art. 36 of the Istanbul Convention.

violence.[80] According to the Court, the lack of consent corresponds most directly to the protection of the sexual autonomy of individuals, which is seen as the central value in prohibiting sexual violence. However, although the Court instructed states to focus their definition of rape on non-consent, it did not direct states in the phrasing of domestic provisions, where states were left with a margin of appreciation. The formulation of the elements of rape in domestic criminal law was thus deemed less important than the interpretation given to such concepts. States are as such able to retain a definition requiring the use of force, if force is interpreted broadly to cover non-consensual acts. Such an approach has also been adopted in relation to the provision on sexual violence in the Istanbul Convention.[81]

The definition of rape has not been analysed as extensively by other regional or international bodies. Nevertheless, in the case of *Miguel Castro-Castro Prison v Peru* in 2006, the IACtHR likewise considered both international criminal law and comparative domestic criminal law and emphasised the element of non-consent, and that a rape victim does not need to demonstrate physical injuries nor that he/she physically resisted the attack.[82] Similarly, the CEDAW Committee in *Vertido v the Philippines* and *R.P.B. v the Philippines* criticised the fact that the domestic penal code in question did not place the lack of consent of the victim at the centre of the definition and recommended the removal of the requirement that sexual assault be committed by force or violence.[83] Meanwhile, international criminal law displays a wider divergence, ranging from a focus on non-consent to coercion or the use of force. [84]

As for the *actus reus* of rape, few regional human rights law courts and UN treaty bodies have expounded on the type of acts included. Commonly, a narrow approach to the *actus reus* has been taken at the domestic level, focusing on heterosexual intercourse. This stems from the historical male control of female sexuality, with the

[80] *M.C. v Bulgaria* (ECtHR). In this case, the applicant complained to the ECtHR that Bulgarian law did not offer effective protection against sexual violence, since only cases where victims physically resisted rape were prosecuted.

[81] CoE, 'Explanatory Report to the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence' (ETS No. 210) 11 May 2011, para. 191.

[82] *Miguel Castro-Castro Prison v Peru* (IACtHR). The focus on non-consent was also affirmed in *Fernandez Ortega* et al. *v Mexico* (IACtHR) and *Ana, Beatriz and Celia Gonzalez Perez v Mexico*, IACmHR, Case 11.565, Report No 53/01, OEA/Ser.L/V/II.111 Doc. 20 rev. at 1097 (4 April 2001). Rather than force being the essential element of rape, rape was defined as a sexual act perpetrated through the negation of consent, for example, through coercion.

[83] *Vertido v the Philippines* (CEDAW), para. 8.7; *R.P.B. v. The Philippines* (CEDAW), para. 8.10. *See* also the element of non-consent in Art. 5 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[84] *Prosecutor v Jean-Paul Akayesu* (CTR), para. 688: (coercive circumstances); *Prosecutor v Furundzija,* ICTY, Case No. IT-95-17/1-T (10 December 1998), para. 185 (coercion or force); *Prosecutor v Dragoljub Kunarac, Radomir Kovač and Zoran Vuković*, ICTY, Case No. IT-96-23-T & IT-96-23/1-T (22 February 2001), para. 460 (non-consent); Art. 7 (1) (g)-1; Art. 8 (2) (b) (xxii)-1 of the International Criminal Court (ICC), Elements of Crimes (2011) (coercion or force).

risk of pregnancy outside of wedlock as the main concern.[85] At the international level, there is a clear indication that the traditional focus on penile penetration has been abandoned for a broader, gender-neutral approach to invasions of individual autonomy, including vaginal penetration with fingers,[86] the use of objects such as a truncheon[87] and anal sex.[88] The Istanbul Convention similarly defines rape as '[e]ngaging in non-consensual vaginal, anal or oral penetration of the body of another person with any bodily part or object'.[89] The CEDAW Committee in *Vertido v the Philippines* held that states should remove any requirement of proof of penetration.[90] Furthermore, the complaint in *V.P.P. v Bulgaria* before the Committee concerned incidents where the perpetrator had inserted his finger into a child's anus and tried, without succeeding, to insert his penis into the child's vagina.[91] He was prosecuted for the crime of sexual molestation. The CEDAW Committee criticised the fact that the charge was not rape or attempted rape, given that it involved anal penetration of a sexual nature by a body part of the perpetrator, as well as attempted vaginal penetration.[92] In doing so, it referred to the Istanbul Convention.[93] More detailed delineations of the *actus reus* are provided in international criminal law, in consideration of the principle of specificity required in this area of law. The precision of elements has also been deemed necessary in order to distinguish rape from other forms of sexual violence, in order to maintain the gravity of the crime. These definitions also employ a gender-neutral approach but beyond this premise vary as to the constituent acts, nevertheless with a focus on penetration in more recent sources.[94]

---

[85] Dripps (1992), p. 1780.

[86] The IACtHR in *Miguel Castro-Castro Prison v Peru*, involving digital penetration, specifically held that: (. . .) sexual rape does not necessarily imply an non-consensual sexual vaginal relationship, as traditionally considered. Sexual rape must also be understood as act of vaginal or anal penetration, without the victim's consent, through the use of other parts of the aggressor's body or objects, as well as oral penetration with the virile member'. *See Castro-Castro Prison v Peru* (IACtHR), para. 310. *See* also ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3.(1) (b): 'rape which includes penetration of the vagina, anus or mouth by any object or part of the body'.

[87] *Zontul v Greece* App no 12294/07 (ECtHR, 17 January 2012).

[88] *Miguel Castro-Castro Prison v Peru* (IACtHR), para. 310; *V.P.P. v Bulgaria*, CEDAW Communication No. 31/2011, UN Doc. CEDAW/C/53/D/31/2011 (24 November 2012), para. 9.5; *Zontul v Greece* (ECtHR).

[89] Art. 27 of the Istanbul Convention. Cf Art. 5 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'. Although also encompassing vaginal, anal or oral penetration with a body part or object, it is not gender-neutral, focusing on female victims.

[90] *Vertido v the Philippines* (CEDAW), para. 8.9 (ii).

[91] *V.P.P. v Bulgaria* (CEDAW).

[92] ibid., para. 9.5.

[93] ibid., para. 9.5.

[94] The ICTY in the cases of *Kunarac* and *Furundzija* defined rape as penetration of the vagina or anus by genitals or objects or oral penetration by a penis. *See Prosecutor v Furundzija,* ICTY, Case

Furthermore, rape may transpire even if the perpetrator has not physically participated in the sexual act, in instances where two victims are forced to perform sexual acts on each other. Few mentions are made in international human rights law of such offences. Although it is not explicitly defined as an act of rape but encompassed in the spectrum of sexual violence, the Istanbul Convention obliges states to criminalise 'causing another person to engage in non-consensual acts of a sexual nature with a third person'.[95] However, the same definition is included in the concept of rape in the 2022 EU proposal for a directive on violence against women.[96] Meanwhile, the ACmHPR has categorised it as 'compelled rape, i.e. committed by a third person compelled to carry out the abuse'.[97] Cases involving such acts have also been tried in international criminal law. Several cases adjudicated by the International Criminal Tribunal for the Former Yugoslavia (ICTY) and the Special Court for Sierra Leone involve forcing victims to perform sexual acts on each other, including fellatio, categorised as rape.[98] As noted above, the context in which crimes occur in international criminal law may affect the assessment of such elements as "non-consent" or "coercion", however, a distinction is not warranted in relation to the *actus reus*.

Meanwhile, even though forced masturbation to a limited extent has been addressed as a crime in international criminal law, it is unlikely that it would fulfil the criteria of rape, given the current approach to the *actus reus* of the crime involving penetration.[99] Similarly, although the ACmHPR may denote 'any other forced touching that the victim is compelled to perform on himself/herself or a third person' as sexual violence, it has not been categorised as a form of rape.[100] This has

No. IT-95-17/1-T (10 December 1998), para. 185; *Prosecutor v Dragoljub Kunarac, Radomir Kovač and Zoran Vuković*, ICTY, Case No. IT-96-23-T & IT-96-23/1-T (22 February 2001), para. 460. Meanwhile, a broader definition was developed by the ICTR, which in the *Akayesu* case defined rape as a physical invasion of a sexual nature. *See Prosecutor v Jean-Paul Akayesu* (ICTR), para. 688. However, subsequent case law of the ICTR has adopted the approach of the ICTY, that is, the insistence on clearly defined acts. *See Prosecutor v Mikaeli Muhimana*, ICTR, Case No. ICTR-95-1B-T (28 April 2005), para. 550. The Elements of Crimes of the ICC similarly requires penetration of the vagina or anus by either a body part or an object, or oral penetration by a penis. *See* Art. 7 (1) (g)-1, Elements of Crimes. Although this definition, similar to that in *Akayesu*, refers to "invasion", it is narrowed by the requirement of penetration.

[95] Art. 36 of the Istanbul Convention.

[96] Art. 5 (b), European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[97] ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3 (1) (b).

[98] *Prosecutor v Issa Hassan Sesay, Morris Kallon and Augustine Gbao*, Case No. SCSL-04-15-T, Judgment of 2 March 2009, paras. 1191–1195; *Prosecutor v Cesic,* Sentencing Judgment, Case No. IT-95-10/1-S, ICTY, Judgment 11 March 2004, paras. 13–14; *Prosecutor v Delalic* et al. *(Celebici Camp),* Judgment of 16 November 1998, para. 1066.

[99] *Prosecutor v Milan Martic*, ICTY Case No. IT-95-11-T (12 June 2007), fn 899. *See* also Sivakumaran (2007), p. 267.

[100] ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3 (1) (b).

not been addressed by other sources in international law. Furthermore, coerced online nudity has not been mentioned in international law sources, beyond the online sexual exploitation of children, which is a broader concept.[101] Nevertheless, it appears that forced nudity as a human rights law violation involves coerced nudity without the requisite purpose of a public function, with the aim of sexual or other forms of humiliation. [102] Although often taking place during interrogation, detention or armed conflict, instilling fear, vulnerability and a loss of dignity in the individual, it has been addressed more broadly by a range of international organisations as a form of sexual violence, not restricted to specific actors or settings, within the context of international human rights law.[103] On the Internet, forced nudity often occurs through the coerced undressing in front of a web camera, through sextortion, human trafficking or other means of coercion.[104] Nevertheless, in contrast to rape or forced masturbation, it does not necessarily involve the performance of a physical act, beyond undressing. The lack of penetration entails that it is not considered as grave as certain other forms of sexual violence, such as rape, at the international level.

The categorisation of acts as particular crimes appears to an extent to fall within the margin of appreciation of states, although it may affect certain groups in particular. In *C. A. S. v Romania* of the ECtHR, the complaint concerned in part the fact that the definition of rape through its *actus reus* excluded male victims.[105] Male victims of sexual abuse were only able to file complaints under separate and, arguably, lower status crimes. Nevertheless, the Court held that men were also protected, although through a different provision than rape. The argument that this did not provide as effective protection was refuted by the Court. Since the protection against sexual violence is evaluated from the standpoint of effectiveness in preventing the crime, the main point of concern for the Court is thus whether the act in question is criminalised in some manner. It is, however, possible that the

[101] Art. 23 of the Lanzarote Convention (solicitation of children for sexual purposes).

[102] There is no definition of forced nudity in either international human rights law or international criminal law, such as the contextual elements required. Other terminology may thus be employed to address similar acts, and broader regulations—such as prohibitions on sexual violence—cover such conduct, although not explicitly. *See* case law on forced nudity: *Prosecutor v Jean-Paul Akayesu* (ICTR), para. 697; *Abramova v Belarus* (CEDAW); *Miguel Castro-Castro Prison v Peru* (IACtHR). Being instructed to remove one's clothes may fulfil a legitimate state aim, for example, when searching for contraband in state institutions such as prisons. *See,* for example, *Iwanczuk v Poland* (2001) 38 EHRR 8, para. 59.

[103] Sivakumaran (2007), p. 269; CEDAW, 'Concluding Observations on India' (22 October 2010) UN Doc. CEDAW/C/IND/CO/SP.1, para. 5; ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', para. 58; ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3 (1) (a) and (b).

[104] UNICEF, 'Child Safety Online: Global Challenges and Strategies' (2012) <https://www.unicef-irc.org/publications/pdf/ict_techreport3_eng.pdf> Accessed 16 March 2022, pp. 28, 36, 40.

[105] *C. A. S. v Romania* App no 26692/05 (ECtHR, 20 March 2012).

categorisation itself may lead to a violation, particularly if the punishment attached to the offence is disproportionate to the level of harm. For example, the CEDAW Committee criticised the state in *V.P.P. v Bulgaria* for categorising and prosecuting anal penetration with a finger as sexual molestation rather than rape, which did not reflect the gravity of the offence, in addition to generating a lower sentence.[106] It could also be argued that there is a stronger deterring effect, from a symbolic standpoint, if an act is categorised as "rape" as opposed to "sexual assault", since a greater level of stigma generally is attached to the former.

The approach to the elements of the crimes requires reconsideration in relation to the Internet. This concerns the assessment of non-consent/coercion/force, the *actus reus* of rape and the delineation of modes of liability, for example, direct perpetration or instigation. It is increasingly common that individuals order, watch or broadcast sexual violence digitally. If live-streamed over the Internet, there is no trace of the abuse, creating particular challenges in investigating the offence. Even if a coerced physical sexual act occurs, it may originate from an individual ordering rape or live nudity on a webcam. In such cases, no physical touching occurs between the victim and the perpetrator. The physical act is rather perpetrated by another individual (rape by proxy) or through coerced acts performed by the victim her/himself.

As noted, a focus on non-consent as an element of rape is increasingly required in international human rights law. Such may arise in situations beyond physical force, for example, through verbal threats, rewards, intimidation or some other form of pressure, which may be perceived just as forceful as if it were in person.[107] This element is applicable to online forms of sexual violence. Sextortion, for instance, involves '. . .the use of ICT to blackmail a victim. . .[where] the perpetrator threatens to release intimate pictures of the victim in order to extort additional explicit photos, videos, sexual acts or sex from the victim'.[108] Evaluating non-consent requires a contextual approach regardless of the forum, including the preceding interaction between the alleged victim and perpetrator. Nevertheless, in the context of the Internet, situations may arise where the voluntary nature of, for example, live performances by adult victims must be evaluated, as prostitution and participation in pornography are not explicitly prohibited at the international level and occur in consensual forms online.

As for the *actus reus*, it is clear that the definition of rape at the international level, for example, encompasses acts where the perpetrator has not physically participated, but where physical acts have occurred between two people. Sexual coercion committed through ICTs may in such instances constitute rape. Coerced sexual acts performed by the victim her/himself, even if encompassing penetration as *per* the

[106] *V.P.P. v Bulgaria* (CEDAW), para. 9.5.

[107] Barak (2005), p. 80.

[108] UNHRC, '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), para. 35. *See*, also, definition by GREVIO: CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', p. 31.

definition may, however, not be considered as grave as physical invasions by another individual. It is plausible that such acts are thus categorised as other forms of sexual violence. As mentioned above, it appears that either penetration *of* a sexual organ, or penetration *with* a sexual organ is required at the international level in order to constitute rape. There are no cases at the international level where this has involved acts performed by the victim her/himself. This may *per se* exclude many coerced sexual acts that occur over the Internet. Domestic laws on rape are in many states similarly restricted to coerced physical intercourse or other sexual acts between the perpetrator and victim. For example, forced masturbation has mainly been prosecuted as extortion, sexual coercion or child pornography, but rarely rape, at the domestic level.[109] However, examples also exist where individuals have been convicted of rape over the Internet though never having met the victims, through coercing children to perform sexual acts in front of a webcamera.[110] In such instances, domestic courts have held that the acts performed by the victims themselves were of comparable gravity to physical intercourse, as they involved penetration with fingers or objects. If the focus remains on the harm to sexual autonomy rather than the physical invasion by the perpetrator there could, however, be room to argue that the elements of the *actus reus* may also be fulfilled through coerced sexual acts performed by the victim her/himself, regardless of whether this involves penetration or not. However, there is currently no explicit support in international human rights law for this view, beyond a theoretical argument.

Linked to the *actus reus* of the offence, the assessment of the modes of liability is complex in the context of the Internet, where incidents involving perpetrators who initiate but do not physically commit the acts are often charged as "incitement" to rape.[111] Many states exclude such sexual acts from the scope of rape through the *actus reus* of the definition as well as the delineation of individual criminal liability. A distinction is thus made in relation to the online environment connected to the involvement in the physical act. Nevertheless, the extension of concepts of rape are increasingly occurring in domestic law. Again, in order to consider the harm of such acts to the sexual autonomy of victims, the view that perpetrators are less culpable when coercing sexual acts over the Internet undermines effective protection.

[109] Overview of US cases in Brookings Institute, 'Closing the sextortion sentencing gap: A legislative proposal' (11 May 2016) <https://www.brookings.edu/research/closing-the-sextortion-sentencing-gap-a-legislative-proposal/#footnote-4> Accessed 16 March 2022.

[110] Svea Hovrätts dom den 16 april 2019 i mål nr B 11734-17 (Sweden); NJA 2015 s. 501 (Sweden); Hovrätten för Västra Sveriges dom i mål B 1082–19 (Sweden).

[111] Hovrätten för Nedre Norrlands dom den 22 maj 2018 i mål nr B 284-18. In the US, the use of webcams for the performance of sexually explicit acts has led to charges for the crime of 'attempted coercion and enticement of a minor' and 'attempted production of child pornography', through 18 U.S.C § 2251 & 2422.

4.2.4.2 Obligations to Investigate and Provide Remedies

Most cases concerning sexual violence at the international level involve state negligence to conduct effective investigations. The regional human rights law courts have developed corresponding standards in this regard. The ECtHR has mainly analysed flaws in domestic investigations through the procedural limb of Article 3. A similar standard of "effectiveness" as *vis-à-vis* criminalisation is required, entailing that investigations should be capable of leading to the establishment of the facts of the case and to the identification and punishment of those responsible.[112] It is thus an obligation of means rather than results. The ECtHR, the IACtHR and the CEDAW Committee have in several cases indicated concrete steps for an investigation to be effective.[113] For example, investigations must be independent, impartial, subject to public scrutiny and diligently performed. A context-sensitive assessment must be made in order to evaluate the credibility of the parties.[114] This is essential in investigations of sexual violence, where witnesses are rare and physical evidence may be lacking.

The Internet as the forum for sexual violence substantially impedes possibilities of conducting effective investigations since perpetrators frequently are anonymous, obligations of Internet intermediaries to aid investigations are limited and international human rights law and EU law place restrictions on data collection and data retention. However, as noted above, the ECtHR in *K.U. v Finland* affirmed that the obligation to effectively investigate online sexual offences included the procurement of data from Internet intermediaries identifying the perpetrator.[115] Although the case involved child sexual abuse, this approach must concern also adult victims, given the severity of sexual violence *per se*. Furthermore, in *Khadija Ismayilova v Azerbaijan* of the ECtHR, the failure of the state to conduct an effective criminal investigation of intimate images published online was considered a violation of Article 8. Specifically, the investigation should have sought to identify the source and uploaders of the videos as well as the owners and/or operators of the websites where the videos were posted.[116]

The IACtHR and the CEDAW Committee in particular have addressed the obligation to investigate sexual violence from a gender perspective, with a focus on the access to justice for victims. Ineffective investigations and gendered

[112] *Aydin v Turkey* (ECtHR), para. 107; *M.C. v Bulgaria* (ECtHR), para, 32.

[113] *See Fernandez Ortega* et al. *v Mexico* (IACtHR), para. 194; *X and Y v Georgia*, CEDAW Communication, CEDAW/C/61/D/24/2009 (25 August 2015); *Maslova and Nalbandov v Russia* (ECtHR), para. 91.

[114] *M.C. v Bulgaria* (ECtHR), paras. 161–163; CoE, 'Explanatory Report to the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence', para. 191.

[115] *K.U. v Finland* (ECtHR).

[116] ibid., para. 127.

assumptions made during investigations have been viewed as gender-based.[117] Similarly, according to the explanatory report to the Istanbul Convention, allowing gender stereotypes on female and male sexuality to influence investigations on sexual violence contravenes the Convention.[118] According to the UN HRC, questioning the morality of the rape victim during the investigation, such as inferring promiscuity, is discriminatory and contributes to re-victimisation.[119] Additionally, examining the sexual life of victims during the course of investigations of rape is categorised as a violation of the right to privacy *per se*. [120] This is particularly relevant in instances of image-based sexual abuse where the victim has consented to the taking of the photograph. As noted above, the harm is often trivialised, with the morality of the victim called into question by authorities. Similarly, gender-stereotyped interpretations of coercion, as solely involving physical force, undermine effective investigations of online sexual abuse where coercion is mainly speech-based. Furthermore, negligence to investigate is often linked to the attitude that sexuality and social relationships belong to the private sphere, free from state intrusion, which enforces the notion that women are subordinate and thus not worthy of equal protection. With due regard of these paradigmatic problems, the IACtHR has held that in cases of violence against women, states have "enhanced" due diligence obligations, entailing that the above requirements are particularly urgent and extensive.[121] The IACtHR has also affirmed certain minimum standards to ensure that investigations are conducted in a gender-sensitive manner.[122]

Furthermore, as made clear in *O'Keeffe v Ireland* of the ECtHR, states must establish effective mechanisms for the detection and reporting of sexual violence, with the requirement heightened in contexts generating particular risks of sexual abuse, in this case denominational schools during the particular time span reviewed in the case.[123] According to the Court, a remedy is deemed to be ineffective if the ill-conduct continues during a long period of time, indicating that the state ought to have been aware of the problem. Through such reasoning, the widespread occurrence of a violation is an indication of the inadequacy of measures adopted by the

[117] A failure to effectively investigate rape has been categorised by the IACtHR/IACmHR as sex discrimination. *See*, for example, *Ana, Beatriz and Celia Gonzalez Perez v Mexico* (IACmHR); *Fernandez Ortega* et al. *v Mexico* (IACtHR). *See* also *Case Egyptian Initiative for Personal Rights and Interights v Egypt* (ACmHPR), para. 179; *L.N.P. v Argentina* (UNHRC), para. 13.3; *X and Y v. Georgia* (CEDAW), para. 9.7.

[118] CoE, 'Explanatory Report to the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence', para. 192.

[119] *L.N.P. v Argentina* (UNHRC), para. 13.3.

[120] ibid., para. 13.7.

[121] The general obligations in the American Convention are complemented by the obligations in the Convention of Belém do Pará. *See*, for example, *Fernandez Ortega* et al. *v Mexico* (IACtHR), para. 193.

[122] This includes authorities taking statements in private; prompt medical and psychological assessments; continuous psychological treatment and health-care; and free legal assistance.

[123] *O'Keeffe v Ireland* (2015) 59 EHRR 15.

state, which implies an obligation of results rather than solely of means. The CEDAW Committee has also held that a right to a remedy is implied in the Convention, primarily in Article 2 (c), which requires that states ensure effective protection of women against discrimination through competent domestic courts.[124] Effective remedies include conducting proceedings in a fair, impartial, timely and expeditious manner,[125] sufficiently severe punishment to ensure deterrence[126] and compensation for the harm suffered.[127] Given the widespread occurrence of online harassment, domestic remedies can from this standpoint be categorised as ineffective. Nevertheless, this mainly stems from the lack of criminalisation of the acts, ineffective investigations by law enforcement and the transboundary nature of online communication, causing jurisdictional complexities, for example, in the enforcement of domestic judgments.

Furthermore, states must undertake operational measures in certain situations in order to fulfil their obligations to protect, according to the ECtHR, where the state 'knew or ought to have known' that a violation would occur.[128] Most commonly, such obligations arise in situations of repeated abuse[129] but could also involve instances of rape threats. Finally, the Inter-American Commission on Human Rights (IACmHR), the IACtHR and the CEDAW Committee have, due the pervasive and gendered nature of sexual violence, recognised that broader measures aiming to prevent violations are necessary. For example, states are obliged to adopt coherent prevention strategies, to analyse and reduce risk factors such as gender inequality and to strengthen institutions to effectively prevent and respond to instances of sexual violence, including training public officials.[130]

[124] *Vertido v the Philippines* (CEDAW), para. 8.3. *See* also *Fulmati Nyaya v Nepal* (UN HRC), para. 9.

[125] *Vertido v the Philippines* (CEDAW), para. 8.3; *R.P.B. v. The Philippines* (CEDAW), para. 8.2.

[126] *V.P.P. v Bulgaria* (CEDAW), para. 9.5; *Zontul v Greece* (ECtHR); Art. 45 of the Istanbul Convention. As noted by the UN Special Rapporteur on Violence against women, effective sanctions are necessary to prevent future conduct of a similar nature. Penalties must thus be sufficiently severe. *See* UNHRC, 'Report of the Special Rapporteur on Violence against Women, its causes and consequences, Rashida Manjoo' (14 May 2013) UN Doc. A/HRC/23/49, para. 74.

[127] *V.P.P. v Bulgaria* (CEDAW), para. 9.11.

[128] *Osman v the United Kingdom* (2000) 29 EHRR 245, para. 116. Accordingly, '[a] failure to take reasonably available measures which could have had a real prospect of altering the outcome or mitigating the harm is sufficient to engage the responsibility of the State'. *See E. and Others v the United Kingdom* (2003) 36 EHRR 31, para. 99.

[129] *E. and Others v the United Kingdom* (ECtHR). The authorities had failed to remove children from a home despite strong indications of continued abuse.

[130] *González* et al. *(Cotton Field) v Mexico* (preliminary objections, merits, reparations and costs) IACtHR Series C No. 205 (16 November 2009); *X and Y v Georgia* (CEDAW), para. 9.11.

4.2.4.3 Regulating Intermediary Liability

As viewed, international human rights law has in relation to the obligation to protect focused on the liability of the perpetrator of sexual violence, be it a state or non-state actor. This approach remains online. Meanwhile, Internet intermediaries provide the platforms through which sexual violence is perpetrated. Although these do not incur legal obligations to protect individuals, state obligations to adopt domestic laws or procedures that involve intermediaries may arise, as evident in *K.U. v Finland*. Nevertheless, while the case law of the ECtHR has affirmed obligations to monitor and remove hate speech—as a consequence of both the profound harm and the purported ease with which such offences can be detected—such obligations have not as of yet been extended to sexual violence. Furthermore, online sexual violence has not been explicitly affirmed by the EU as illegal conduct in terms of intermediary responsibility. Although the proposed EU directive on violence against women includes obligations to criminalise rape, it does not involve the regulation of intermediary liability in this regard, solely *vis-à-vis* "cyber violence".[131] Given that sexual violence is considered among the most severe forms of human rights law violations, it is likely a result of a lack of harmonisation on the scope of technology-facilitated sexual abuse as well as a pragmatic approach to the ability of intermediaries to control and assess this type of offence.

In considering the potential development of state obligations to regulate intermediaries, the function and characteristics of the particular intermediary affects the means of prevention and thus the content of obligations. The scope of protection is more limited in relation to private online pockets as opposed to conduct on public fora. That is, conversations and conduct through private email, chat functions and live-streaming are generally beyond the purview of intermediary regulation. In most instances, ICTs are mere facilitators of social contact, that is, meeting places similar to any public space. They are not responsible for the social engagement between individuals that leads to sexual violence, for example, through dating websites. Nevertheless, intermediaries and website operators facilitating personal contact, whether sexual or not, are increasingly adopting certain security measures, for example, report abuse functions or identity verification systems on dating sites.[132] While this may not be construed as legal obligations, it can be discussed in terms of soft law recommendations, in accordance with norms on corporate social responsibility.

Where the characteristics of platforms entail a high risk of severe sexual abuse and the level of control of content is more extensive, the potential for stricter requirements of monitoring should, however, be explored. This must be balanced

[131] Art. 25 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[132] *See*, for example, Euronews, 'Tinder to roll out ID verification worldwide in bid to make online dating more "safe"' (17 August 2021) <https://www.euronews.com/next/2021/08/17/tinder-to-roll-out-id-verification-worldwide-in-bid-to-make-online-dating-more-safe> Accessed 16 March 2022.

against restrictions on general monitoring, for example, in EU law. For instance, stricter liability has been discussed in relation to websites encouraging revenge pornography, through their format and structure.[133] Such intermediaries are the most suitable points of control where images are accessible to the public.[134] As this has yet to be addressed in international human rights law, it has not been construed as an obligation for states. Given that positive obligations have been formulated in relation to the protection against the non-consensual publication of photographs in the media, its extension to secondary liability for media publishers and, plausibly, also intermediaries, would most likely involve notice-and-takedown systems. For example, Twitter and Facebook have flagging systems for graphic content, such as pornography, overt sexual content, violence or other illegal activities, where users may notify the company of illicit content.

However, notification-based systems involve a moderating process, which often takes days to weeks and poses a particular problem for live videos. Certain video platforms, such as Periscope or Facebook, transmit live content. No recording is saved on the website and companies often do not monitor posts. This is thus an area that requires technological development and where law may guide design. The notice-and-takedown system may, for example, take specific forms. As noted above, in relation to platforms allowing live feed, it has been suggested that "urgent reaction" buttons must be included, for immediate human moderation, removal and contact with authorities in instances of illegal conduct, including sexual assault. The development of AI-based systems of notification is also underway, for example, in relation to image-based sexual abuse. The question nevertheless arises whether users, human moderators and AI possess the knowledge and ability to assess contextual elements such as "coercion" in relation to adults, although regulated by internal guidelines. In terms of image-based sexual abuse, it is generally beyond the capabilities of intermediaries to determine whether an image has been recorded or taken with consent. The same concerns live feed. Since most forms of pornography are considered lawful, often performed by amateurs online, it is particularly challenging. These practical concerns, as well as rule of law considerations, may be the reasons that the proposed EU directive on violence against women includes obligations to ensure the removal or disabling of access to materials of image-based sexual abuse through domestic judicial orders, rather than through other forms of notices.[135]

Remedies when intimate images have been published also include the right to be forgotten in EU law. Additionally, according to the ECtHR, restricting the redistribution of material that may cause injury to individual integrity is in certain instances warranted in order to protect privacy.[136] As argued by McGlynn and Rackley,

[133] Citron (2014), p. 175; Patton (2015), p. 408.

[134] Suzor et al. (2017), p. 1066.

[135] Art. 25 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[136] *Aleksey Ovchinnikov v Russia* App no 24061/04 (ECtHR, 16 December 2010).

regulating secondary distribution is essential as the Internet enables instant and widespread dissemination, considerably intensifying the harm of victims.[137] Incentives by Internet intermediaries to address the issue have been taken also in this regard. For example, several search engines (Google, Bing & Yahoo) provide webforms for requesting the removal of non-consensually uploaded nude or sexually explicit material. Meanwhile, Facebook has introduced new photo-matching technology to prevent photographs from being redistributed.[138] Nevertheless, a more developed and comprehensive approach to the positive obligations of states *vis-à-vis* intermediaries is warranted, given the gravity of sexual violence.

It should be noted that simulated sexual violence is at times raised as a violation of individual sexual autonomy, where a person's digital representation of themselves, for example, an avatar, is subjected to sexual violence, most frequently in gaming worlds. This may result in emotional harm, especially if the individual has previously been a victim of assault.[139] It has been discussed as a form of "virtual rape", but it is not considered to involve the same level of harm as that of physical abuse.[140] Rather, regulation of ICTs that provide such products may be addressed as a means of eliminating gender stereotypes, for example, as state obligations under Article 5 of the CEDAW.

4.2.5 Conclusions

The prohibition of sexual violence, whether committed by state or non-state actors, is encompassed in various areas of public international law. While an aspect of several rights, its core aim is to protect sexual autonomy, human dignity and physical integrity, for example, advanced through the right to privacy and the prohibition of torture, inhuman or degrading treatment. Additionally, sexual violence has been considered gender-based *per se*, particularly in relation to women as a group, given its causes, nature and disproportionate impact on women. The CEDAW Committee, the UN HRC, the Inter-American human rights institutions and the ACmHPR have thus held that it either constitutes gender discrimination in general or under particular circumstances. Since sexual violence, in addition to the harm and fear it instils, may cause women to retreat from the Internet, the prevention and remedying of such violence on the Internet is thus a necessary factor in ensuring gender equality in terms of both access to and safety on the Internet.

---

[137] McGlynn and Rackley (2017), p. 538.

[138] Facebook, 'Detecting Non-Consensual Intimate Images and Supporting Victims' (15 March 2019) <https://about.fb.com/news/2019/03/detecting-non-consensual-intimate-images/> Accessed 15 March 2022.

[139] Powell and Henry (2017), p. 92.

[140] The Guardian, 'Sexual harassment in virtual reality feels all too real – 'it's creepy beyond creepy'' (26 October 2016) <https://www.theguardian.com/technology/2016/oct/26/virtual-reality-sexual-harassment-online-groping-quivr> Accessed 16 March 2022.

The protection of individuals against sexual violence is a state obligation regardless of the medium, for instance, recognised by the ECtHR in *K.U. v Finland*. At a cursory level, there is thus no gap in regulation and an equivalence exists between protection IRL and online. However, the medium affects certain legal assessments, such as the categorisation of the act in question as a form of sexual violence *per se* or a particular type of sexual abuse, such as rape. The degree of physical interaction additionally affects the assessment of the severity of the harm and thus the applicability of specific provisions in international human rights law. The main challenge in transposing current standards on sexual violence is thus the approach to harm. These concepts require both context- and gender-sensitive interpretations, recognising means of coercing sexual acts through new technology. The online context also affects the content of state obligations, such as the means of conducting effective investigations and regulating intermediary liability.

It is necessary as a first step to categorise various forms of sexual coercion online—both physical acts and speech-based conduct—as sexual violence and, secondly, to consider how these correspond to the definitions of the various offences. The harm of sexual violence to the individual is in international sources generally described as physical and psychological. Whereas the aim of protecting sexual autonomy remains the same online/offline, certain offences commonly require physical injury. It has been argued that legal norms concerning physical acts cannot be transposed to cyberspace. Accordingly, '[l]aws which address the mental states of actors, or the outcomes of their behaviours, tend to require little adaptation to achieve equivalence. In contrast, laws which address the ways in which behaviour is undertaken can pose difficult challenges for lawmakers, because the cyberspace technologies tend to incentivize actors to undertake those behaviours very differently'.[141] For example, it is easier to transpose criminal laws on stalking and harassment as the elements of repetitive, intrusive behaviour remain the same when committed through ICTs. The same may not be the case for regulations defining offences in terms of specific physical acts of sexual violence.

It is clear that rape, forced masturbation and forced nudity are considered forms of sexual violence at the international level, engaging positive obligations for states to protect individuals, including the adoption of effective criminal laws and effectively investigating offences. However, since most offences on the Internet are conducted through speech—textual or verbal—there is a risk that the act/harm is trivialised as "mere speech", excluding certain sexual interactions from the scope of "sexual violence". For example, threats of physical or sexual violence are frequently considered dignitary harms, rather than forms of sexual violence *per se.* Similarly, sexual harassment is generally considered a form of gender-based violence in international human rights law but not sexual violence, unless involving physical acts. Meanwhile, image-based sexual abuse—particularly the non-consensual distribution of intimate images—is increasingly categorised as sexual violence, although not comprehensively in the context of international human rights law.

[141] Reed (2012), p. 109.

For example, although the ECtHR in *Söderman v Sweden* and *Khadija Ismayilova v Azerbaijan* affirmed obligations to protect the psychological integrity of individuals, including sexual autonomy, filming a person in the nude and disseminating intimate images online were not explicitly categorised as forms of sexual violence. The characterising of speech as sexual violence is not solely relevant from a symbolic standpoint but also in terms of the content of state obligations. The ECtHR has held that regulation through civil law is not sufficient to protect individuals against sexual violence. However, it is deemed adequate in relation to offences such as defamation, intentional infliction of emotional distress and the non-consensual publication of private information. Where harm is speech-based, the reluctance to restrict the freedom of expression also has an effect on regulation.

In international human rights law, there is a hierarchy attached to the gravity of the harm, distinguishing physical from non-physical acts of sexual violence as well as penetration from other less invasive physical acts. For instance, while rape perpetrated by a state actor has been considered a form of torture, and rape by non-state actors as inhuman or degrading treatment and an invasion of privacy, non-physical acts of sexual violence, such as filming a person in the nude or forced nudity, have been categorised as violations primarily of the right to privacy. A distinction is also noticeable in international criminal law, with rape considered more severe than forced nudity and forced masturbation. The categorisation of the underlying acts of sexual violence as a consequence matters, affecting which human rights norms are applicable and the content and scope of state obligations. Thus, whether coerced nudity or sexual acts in front of a web camera are characterised as rape, sexual molestation, sexual harassment or forced nudity holds legal significance in international human rights law. While the structure of state obligations remains the same in terms of protection—that is, criminalisation, investigation, prosecution, reparations and operational remedies—the content is more extensive *vis-à-vis* rape. The classification of the offence also affects the level of stigma and thus, arguably, the level of deterrence.

Several aspects of the legal approach to the criminal elements of, particularly, rape place an obstacle to recognising this form of sexual violence in the digital context. Rape conducted through the Internet does not generally involve the use of physical force, but rather coercion through speech, for example, in the form of threats. According to international human rights law, states must adopt "effective" criminal laws prohibiting rape. In this regard, an emphasis has been placed on non-consent as the integral element in the definition of rape, which encompasses other forms of coercion than physical force and would thus recognise instances of sextortion. Consent is also central in relation to other forms of sexual violence. The underlying act is not unlawful in adult interactions if consent is present. This element thus translates to the online context. However, a gender-sensitive approach in assessing non-consent is necessary, considering various forms of coercion online, as pressure may be put on the victim, for example, in the context of sextortion or domestic violence.

Meanwhile, the approach to the *actus reus* of rape is not as coherent at the international level. Nevertheless, sources in ICL, EU law and international human

rights law indicate that although penetration or invasion is generally required, this need not involve physical acts committed by the perpetrator. It also encompasses situations where two victims are forced to perform sexual acts on each other. No international cases involve coerced sexual acts by the victim him−/herself. Such an approach has, however, been taken by certain domestic courts in instances where the physical act involves penetration, given that it entails both psychological and physical injury. Meanwhile, although limited sources indicate that forced masturbation may be a violation of international human rights and international criminal law as a form of sexual violence, it appears to be distinguished from rape on the basis of a lack of penetration. A similar approach has been taken in relation to forced nudity.

Given the *actus reus* of various forms of sexual violence and the hierarchy constructed *vis-à-vis* such violence in international human rights law, offences of a sexual nature through digital means are thus generally excluded from the scope of certain crimes, such as rape. This entails that acts of sexual violence on the Internet are categorised as being of lesser gravity than those IRL. The protection of sexual autonomy is accordingly construed narrowly to encompass solely physical integrity, involving certain physical acts. The lack of a physical component should, however, not entail that the harm is viewed as minor. Retaining the focus on harm to sexual autonomy entails that the purpose and aim of laws remain the same, taking the technical environment into account. Many feminist legal scholars have in fact challenged domestic laws on rape requiring vaginal penetration, asserting that sexual autonomy is harmed also through other forms of sexual acts.[142] An extension of the *actus reus* is thus necessary so that it includes a wider scope of physical acts that violate the sexual autonomy of the individual, for instance, coerced penetration performed by the victim him/herself.

As the perpetrator of digital sexual violence generally does not engage physically, it also presents challenges in terms of determining liability. Domestic courts struggle with delineating responsibility, which is often construed as "incitement", "aiding or abetting" or "criminal conspiracy" in instances where the perpetrator has coerced or ordered acts but not physically participated. Although the issue is generally considered to be within the authority of domestic legislation, and thus beyond the involvement of international human rights law, it is possible to argue that a focus on the harm to the sexual autonomy of the victim and the causality between the perpetrator and the act requires the application of direct individual responsibility, in order to ensure effective protection against sexual violence in the digital sphere. It is also linked to the definition of the various acts of sexual violence.

Beyond the adoption of criminal laws encompassing online forms of sexual abuse, obligations also include conducting effective investigations in a gender-sensitive and context-specific manner. Specific standards have been developed by

[142] For an overview on various feminist views, *see* Quenivet (2005), p. 13. *See* also Temkin (2002), p. 57. This may, for example, also involve the use of fingers in the vagina or anus, or a sexual organ penetrating another orifice—such as the mouth. A narrow *actus reus* that focuses on vaginal penetration also only allows for female victims and male perpetrators, in contrast to a liberal approach.

a range of human rights law bodies. Effective investigations of sexual violence require inquiries void of moral considerations of the sexual life of the victim, while bearing in mind new sexual practices online. This is, for instance, relevant in cases of the non-consensual publications of intimate images, where the victim has consensually taken and/or sent the image, as well as in cases of sextortion. Consent must be assessed in relation to the specific act in question and not be presumed, for example, on the basis of consent to record an image in the context of image-based sexual abuse. Naturally, the harm is not reduced merely by a person having exercised his/her sexual autonomy in this manner. Although a right to anonymity exists, the identification of perpetrators of these forms of violence is a necessary aspect of effectively investigating offences. The ECtHR in *K.U. v Finland* affirmed an obligation for states to adopt a legal framework that allows for the identification of perpetrators, on this limited issue, thus affirming that the same obligations apply online. Nevertheless, substantial practical challenges remain in identifying perpetrators, which increases the incentive to regulate the liability of intermediaries.

Sources indicating an approach to intermediary liability *vis-à-vis* sexual violence are scarce. Whereas the majority of states criminalise some forms of sexual violence, the scope and content of such differ greatly. Thus, whereas such acts are illegal, rather than merely harmful, the lack of international harmonisation has undermined effective regulation of such acts online, including intermediary liability. Additional factors impede this development. Despite the severity of the harm, regulation would most likely be balanced against such practical concerns as the ability of intermediaries to assess the legality of acts. Sexual violence frequently occurs through end-to-end activity, in private pockets where intermediaries have limited control over the environment. In view of the general approach to offences other than hate speech—such as disclosure of private information in the media and online defamation—it is likely that positive obligations to protect through regulation will be affirmed but in the form of notice-and-takedown mechanisms. Whereas arguments have been presented for additional components to the notice-and-takedown system, for example, in relation to revenge pornography websites and platforms offering live feed, such are currently not construed as obligations. However, a plethora of options are available for states, in conjunction with Internet intermediaries, to develop means of detecting and reporting sexual abuse. Given the severity of sexual violence, a comprehensive and concrete approach is necessary.

4.3 Harassment

4.3.1 Introduction

"Online harassment" has not been categorically defined in international human rights law. Whereas certain soft law sources approach it in terms of the intentional infliction of substantial distress through online speech, sufficiently persistent to

amount to a course of conduct,[143] the inclusion of isolated incidents is increasingly recognised.[144] For example, the definition of sexual harassment in the Istanbul Convention does not limit it to situations of continuous conduct.[145] Depending on the gravity of the act, it is thus plausible that singular incidents could be included. Meanwhile, behaviour that in itself would not reach the requisite threshold of severity in domestic or international law may cumulatively constitute harassment. Notably, the proposed harmonisation of the concept of "cyber harassment" in the EU requires a cumulative attack against another person by making threatening or insulting material accessible to a multitude of end-users online.[146] Nevertheless, the other mentioned sources also include interpersonal conduct. It may involve sending unwanted sexually explicit or intimidating messages, offensive advances on social media or chat rooms, verbal violence and threats of physical, psychological or sexual violence.[147] It may also include the dissemination of defamatory statements or the posting of personal information.[148] Online harassment may additionally comprise cyber stalking, which is a more narrow set of actions and is generally considered a course of conduct that causes, or might cause, a reasonable person to fear for his or her safety.[149] Harassment also overlaps with sexist hate speech, discussed in Sect. 4.4. Harassment is thus a broad concept covering a range of offences, depending on the content and context.[150] As such, "cyber harassment" constitutes a wider concept than "online sexual harassment". Meanwhile, cyber harassment of women commonly takes the form of sexual humiliation or

[143] For example, CoE (2018a), 'Mapping study on cyberviolence', Cybercrime Convention Committee, Working Group on cyberbullying and other forms of online violence, especially against women and children (CBG) <https://rm.coe.int/t-cy-mapping-study-on-cyberviolence-final/1680a1307c> Accessed 7 March 2022, p. 6.

[144] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 38; UNHRC, '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), paras. 38–40; Art. 9 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'; European Institute for Gender Equality (EIGE), 'Cyber Violence against Women and Girls' (2017), p. 2.

[145] Art. 40 of the Istanbul Convention.

[146] Art. 9 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[147] European Institute for Gender Equality (EIGE), 'Cyber Violence against Women and Girls' (2017), p. 2.

[148] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 38.

[149] Cyber-stalking can be accomplished through similar means and achieve similar ends as harassment. It includes sending of messages that are offensive or threatening, posting offensive comments about the respondent on the Internet and sharing intimate photos or videos of the victim on the Internet or by mobile phone, repeatedly and by the same person. *See* European Union Agency for Fundamental Rights, 'Violence against Women: An EU-Wide Survey: Main Results' (2014), p. 87.

[150] Citron (2014).

coercion.[151] Harassment of a sexual nature is also considered particularly harmful, since sexuality is central to individual identity.[152] Meanwhile, the following sections will discuss select forms of cyber harassment, with a focus on sexual harassment, defamation and the non-consensual disclosure of private information.

4.3.2 Sexual Harassment

4.3.2.1 Introduction

The lack of a comprehensive definition of online sexual harassment at the international level entails that empirical studies on the subject are limited. Nevertheless, in a survey undertaken by Anastasia Powell and Nicola Henry—in accordance with their own definition of the offence—there is a clear indication that mainly women, particularly young adults, are subjected to online sexual harassment. The most common form is the unwanted receipt of sexually explicit images, comments or emails and, secondly, repeated and/or unwanted sexual requests online or via email/text messages.[153] These acts are frequently categorised as a variety of civil or criminal law offences at the domestic level, such as the intentional infliction of emotional distress, defamation, indecent exposure or sexual molestation. In contrast, approaching the matter coherently from the viewpoint of "sexual harassment" entails that the gendered causes and consequences—such as gender inequality—are more fully recognised. Such a categorisation places obligations on states to undertake broader measures to combat gender inequality. Additionally, if instances of harassment are solely categorised as defamation or invasions of privacy, certain acts fall outside the scope of protection.[154] Acknowledging the sexual content of harassment also strengthens the argument for including it within the broader category of sexual violence. Furthermore, considering cumulative acts as harassment acknowledges harm that, as a single incident, may not reach the requisite threshold of severity of other forms of violations.

4.3.2.2 From Domestic to International Law

The approach to sexual harassment mainly developed at the domestic level in the 1970s and 1980s, with feminist legal scholarship categorising it as a form of sex discrimination and defining its scope in terms of subjects, acts and contexts.[155] This

[151] Powell and Henry (2017), p. 212.

[152] Chaffin (2008), p. 780.

[153] *See* Powell and Henry (2017), p. 158. *See* also, for example, Barak (2005), p. 78.

[154] Franks (2012), p. 693.

[155] MacKinnon (1979). For an overview, *see* Siegel (2004).

approach in turn influenced international human rights law. Domestic laws on sexual harassment are generally constructed in a contextual manner, focusing on spatial and relational factors, that is, the context in which the sexual harassment has occurred and the relationship between the victim and harasser.[156] The limitation to employment and education settings is linked to several factors. Employment and education are closely connected to the goals of gender equality, primarily understood in liberalist terms. These are considered public places where the effect of sex discrimination is profound.[157] Furthermore, such settings are characterised by a captive audience in contrast to, for example, street harassment, the latter viewed as being something from which you can walk away. Captivity exacerbates the harm.[158] Finally, these environments are generally responsive to effective control.[159]

Domestic law commonly limits sexual harassment provisions to vertical relationships, for example, between superiors and subordinates. This stems from the approach by Catharine MacKinnon that harassment is '. . .the unwanted imposition of sexual requirements in the context of a relationship of unequal power. Central to the concept is the use of power derived from one social sphere to lever benefits or impose deprivations in another'.[160] It is thus a means of producing and maintaining social hierarchies, rather than arising from the individual motivations of the perpetrator. The abuse of power is consequently often deemed a necessary component at the domestic level, in domestic non-discrimination laws or labour laws. Nevertheless, in several states, laws on sexual harassment extend also to *horizontal* relationships and hostile work environments, affirmed also in EU law, including sexual jokes and nude photographs.[161] The hierarchical component is thus not necessarily a vital aspect of the concept.

In contrast, sexual harassment has engendered limited attention within international human rights law, be it through codification, case law or in soft law. An explicit prohibition can be found in regional women's rights conventions. In certain treaties, the approach is broad in terms of settings and the relationship of persons involved. For example, the Istanbul Convention defines it as:

> . . .any form of unwanted verbal, non-verbal or physical conduct of a sexual nature with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment. . .[162]

[156] Franks (2012), p. 662.

[157] ibid., p. 676. Catharine MacKinnon construed it as a form of sexual appropriation of women, inhibiting their equal participation in important public spheres. *See* MacKinnon (1979).

[158] Franks (2012), p. 676.

[159] ibid., p. 675.

[160] MacKinnon (1979), p. 1.

[161] European Parliament resolution on harassment at the workplace (2001/2339(INI)), para, 4. *See*, for instance, in Conte (2010), pp. 11–201.

[162] Art. 40 of the Istanbul Convention. Harassment of a certain nature may also constitute stalking, which is prohibited in Art. 34 of the Istanbul Convention, where it is defined as '. . .intentional conduct of repeatedly engaging in threatening conduct directed at another person, causing her or him to fear for her or his safety.'

This has been affirmed by GREVIO as applying to the Internet.[163] Similarly, Article 2 of the Inter-American Convention on the Prevention, Punishment and Eradication of Violence against Women (the Belém do Pará Convention) provides that sexual harassment in the workplace, educational setting, health facilities, or "any other place" constitutes violence against women.[164] Furthermore, the UN Declaration on Violence against Women categorises sexual harassment in the workplace, educational institutions, and "elsewhere" as a form of violence against women.[165] According to the UN Special Rapporteur on Violence against Women, the definition may vary in accordance with cultural values and norms, but involves sexual attention that is unwanted by, or considered offensive or threatening to the recipient.[166] Consequently, the primary concern is the effect on the victim.

Meanwhile, provisions on sexual harassment in other international sources indicate that the domestic approach to the offence has been influential.[167] The premise that sexual harassment is harmful by limiting the ability of women to participate in the workplace or in education on an equal footing with men has accordingly been transposed.[168] Nevertheless, in certain instances, this has been extended to encompass online harassment. For example, while the CEDAW Committee mainly has discussed sexual harassment in the context of the workplace as a form of inequality in relation to the right to employment,[169] it has also been defined as a form of gender-based violence,[170] with the Committee criticising states in concluding observations for failing to protect women from online harassment.[171] The UN General Assembly

[163] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 36.

[164] The ACmHPR has also held that sexual harassment occurring in the context of a public demonstration was a violation of the African Charter on Human and Peoples' Rights. *See Egyptian Initiative for Personal Rights and Interights v Egypt* (ACmHPR).

[165] Art. 2(b) of the UN Declaration on the Elimination of Violence against Women.

[166] UNCHR, 'Preliminary Report Submitted by the Special Rapporteur on Violence against Women, its Causes and Consequences, Ms. Radhika Coomaraswamy, in Accordance with Commission on Human Rights Resolution 1994/45' (22 November 1994) UN Doc. E/CN.4/1995/42, para. 190.

[167] The CESCR has similarly addressed sexual harassment in relation to just and favourable conditions of work. *See* ICESCR, 'General Comment No. 23 (2016) on the Right to Just and Favourable Conditions of Work (Article 7 of the International Covenant on Economic, Social and Cultural Rights' (27 April 2016) UN Doc. E/C.12/GC/23, para. 48. The Maputo Protocol prohibits sexual harassment in education and employment. *See* Arts. 12(1) (c) and 13(c) (d).

[168] MacKinnon (1979).

[169] CEDAW, 'General Recommendation No. 19: Violence Against Women', paras. 17–18. *See* also Beijing Declaration and Platform for Action, Report of the Fourth World Conference on Women, Beijing, 4–15 September 1995, UN Doc A/CONF.177/20 and UN Doc A/CONF.177/20/Add.1, Platform for Action, para. 178.

[170] CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 14.

[171] CEDAW, 'Concluding Observations on the Combined Sixth and Seventh Periodic Reports of Ireland' (9 March 2017) UN Doc. CEDAW/C/IRL/CO/6–7, para. 26 (c).

in a resolution on sexual harassment addressed this in relation to the workplace, education and 'digital contexts'.[172] Protection against harassment in the workplace is also included in several International Labour Organization (ILO) conventions, including ILO Convention No. 190 from 2019, which extends the concept of the workplace to ICTs.[173] Similarly, EU law—with due regard of its substantive limitations—considers harassment primarily in the context of employment.[174] Nonetheless, the European Parliament has in a resolution called for a reconsideration of the definition of sexual harassment in light of 'social and technological developments and attitudes, which have all evolved and changed over time'.[175] In this regard, it has noted the problem of harassment on the Internet, for example, on social networks, and the blurring of boundaries between private, professional and social life.[176] Accordingly, the Internet is characterised as a public space, with the European Parliament calling for the prohibition of harassment also in this sphere.[177] Thus, cumulatively it appears that international human rights law provides a broader approach to sexual harassment than at the domestic level. Although this increases the demand for state intervention also in the online sphere, as Christine Chinkin notes, a broad approach lacks the specificity of context refined through domestic law and case law.[178] Elements such as the *actus reus* of harassment and liability thus require concretisation if extended beyond the traditional spheres.

Beyond the limited codification of sexual harassment at the international level, the composite acts of harassment may be addressed within general human rights law provisions. Whether sexual harassment consists of acts or offensive speech may in this regard affect the categorisation at the international level and the content of state obligations. Sexual harassment includes both physical and verbal acts. Harassment involving physical violence between private individuals has been deemed to fall

[172] UNGA, 'Intensification of efforts to prevent and eliminate all forms of violence against women and girls: sexual harassment' (14 November 2018) UN Doc. A/C.3/73/L.21/Rev.1.

[173] International Labour Organization (ILO) Discrimination (Employment and Occupation) Convention, 1958 (No. 111); Art. 3(d) of the International Labour Organization (ILO) Violence and Harassment Convention, 2019 (No. 190).

[174] Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services; Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast); European Parliament resolution of 26 October 2017 on combating sexual harassment and abuse in the EU (2017/2897(RSP)).

[175] European Parliament resolution of 11 September 2018 on measures to prevent and combat mobbing and sexual harassment at workplace, in public spaces, and political life in the EU (2018/2055(INI)), para. C.

[176] ibid., paras. K and L.

[177] ibid., general recommendations, para. 48. *See* also European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', which concerns the broader concept of cyber harassment, not specifically sexual harassment online.

[178] Chinkin (2004), p. 655.

within the scope of the prohibition on inhuman or degrading treatment in instances where the state has failed to protect a person, for example, when involving an individual with physical and mental disabilities.[179] This provision also prohibits psychological means of causing harm, such as rape threats, insults and humiliation. However, in general, non-physical violations have been precluded due to lesser gravity.[180] While the ACmHPR in a general comment has considered that harassment, including verbal attacks and humiliation, falls within the scope of the prohibition on torture, inhuman or degrading treatment,[181] other sources thus mainly indicate that online sexual harassment involves violations of the right to privacy.[182] Intrusions of the private sphere include unwanted communication whose purpose or effect it is to undermine the dignity of a person, whether sexual or not. In cases where the content is sexual, it contravenes the sexual autonomy of a person.

As mentioned above, physical acts of sexual harassment can be categorised as forms of sexual violence, while certain soft law sources also include speech-based sexual harassment.[183] The lack of a physical contact does not *per se* exclude it from the realm of sexual violence, evident in the prohibition of forced nudity in international law.[184] Non-physical harassment includes the unsolicited receipt of pornographic or intimate images, which may be considered a form of indecent exposure and thus also sexual violence. Nevertheless, this view is not fully accepted in international human rights law. For example, in the Istanbul Convention, sexual harassment is regulated in a provision separate from sexual violence, despite the definition of harassment encompassing both speech and physical acts.[185] Stricter measures are often required in instances of sexual violence, for example, obliging states to adopt domestic criminal laws. For instance, whereas criminalisation is required in the Istanbul Convention of acts constituting sexual violence, sexual harassment is subject to '. . .criminal or *other legal sanction*', that is, also civil law.[186] It may also affect proportionality assessments and balancing in conflicts of

[179] *Dordevic v Croatia* App no 41526/10 (ECtHR, 24 July 2012).

[180] *Rosa Marta Cerna Alfaro and Ismael Hernández Flores v El Salvador*, IACmHR, Case 10.257, Report No. 10/92, OEA/Ser.L/V/II.81 rev.1 Doc. 6 at 125 (4 February 1992); *Cosans v the United Kingdom* (1982) 4 EHRR 293 (ECtHR).

[181] ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', para. 58.

[182] UN HRC, 'Right to Privacy: Report of the Special Rapporteur on the right to privacy' (16 October 2019), para. 70.

[183] For example, ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3 (1) (b). *See* also the argument in MacKinnon (1993), p. 58, that sexual harassment '. . .is a sexual invasion, an act of sexual aggression, a violation of sexual boundaries, a sex act in itself'.

[184] *See* Sect. 4.2.

[185] Art. 36 and Art. 40 of the Istanbul Convention.

[186] ibid. Emphasis added. Meanwhile, in relation to gender-based violence, states may adopt either criminal or civil legislation, *see*, for example, Art. 7 (c) of the Belém do Pará Convention.

rights. The categorisation of online speech and content as either harassment or sexual violence, for example, in relation to image-based sexual abuse, thus has practical implications.

Furthermore, given that mainly female Internet users are subjected to online harassment, it constitutes a form of sex discrimination. Applying the framework of gender equality is important in order to identify harassment as a group-based harm, rather than solely as an individual offence, and finds support in both feminist legal theory and linguistic philosophy.[187] The CEDAW Committee, the Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (the Maputo Protocol), the Istanbul Convention, the Belém do Pará Convention, the Beijing Platform for Action, the ACmHPR and the European Parliament all consider sexual harassment a form of violence against women and discrimination.[188] Similarly, the ACmHPR in the case of *Egyptian Initiative for Personal Rights and Interights v Egypt* categorised incidents of physical and verbal sexual harassment against female demonstrators in Egypt as forms of gender discrimination.[189] In doing so, it considered the specific incidents and contexts, including the use of derogatory expressions such as "slut" and "whore" directed at women.[190] The Commission noted that women were specifically targeted during the demonstration. The derogatory names used against the women further indicated the gender-related nature of the assaults as they are not commonly used against men and '. . .are generally meant to degrade and rip off the integrity of women who refuse to abide by traditional religious, and even social norms'.[191] In this regard, the Commission considered the acts in the context of the culture in question—an Arab Muslim society—where women's sexuality is associated with purity and honour. However, the use of such words is a common way of demeaning women and may have a similar impact regardless of the cultural context.

The prohibition of sexual harassment in effect involves restrictions on the freedom of expression subject to a contextual assessment, that is, a person is restricted from uttering or writing certain words in particular settings. As noted above, the freedom of expression is a qualified right and it is accordingly subject to legitimate

[187] Franks (2012), p. 696; Fentonmiller (1994), p. 579; MacKinnon (1979). As mentioned in Sect. 3.2.3, there is room to argue that sexually harassing speech involves either perlocutionary or illocutionary speech.

[188] CEDAW, 'General Recommendation No. 19: Violence Against Women', para. 17; Art. 12 (1) (c) and Art. 13 (c) of the Maputo Protocol; Preamble and Art. 40 of the Istanbul Convention; Art. 2(b) of the Belém do Pará Convention; The Beijing Declaration and Platform for Action (Platform for Action), para. 178; ACmHPR, 'General Comment No. 4 on the African Charter on Human and Peoples' Rights: The Right to Redress for Victims of Torture and Other Cruel, Inhuman or Degrading Punishment or Treatment (Article 5)', para. 58; European Parliament resolution of 11 September 2018 on measures to prevent and combat mobbing and sexual harassment at workplace, in public spaces, and political life in the EU (2018/2055(INI)), para. F.

[189] *Egyptian Initiative for Personal Rights and Interights v Egypt* (ACmHPR).

[190] ibid., para. 143.

[191] ibid., para. 143.

state restrictions, including the protection of the rights of others. Whereas certain theories on the freedom of expression embody a broad approach to acceptable speech—such as autonomy-based theories—the freedom of expression does not generally extend to speech vanquishing the autonomy of others, affirmed also by regional human rights law courts. However, challenges arise in identifying what types of speech are considered sexual harassment. As mentioned, sexual harassment is in the Istanbul Convention defined as 'unwanted verbal, non-verbal or physical conduct of a sexual nature with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment', with a similar definition in EU directives on equality in employment.[192] Although analysed in relation to the workplace, the CEDAW Committee has defined sexual harassment as including '. . . such unwelcome sexually determined behaviour as physical contact and advances, sexually coloured remarks, showing pornography and sexual demand, whether by words or actions'.[193] The recent ILO Convention No. 190—also involving the workplace—defines violence and harassment in this context as '. . .a range of unacceptable behaviours and practices, or threats thereof, whether a single occurrence or repeated, that aim at, result in, or are likely to result in physical, psychological, sexual or economic harm, and includes gender-based violence and harassment'.[194]

Nevertheless, the interpretation of these definitions in practice is unclear as cases on this issue have not been brought before international human rights law bodies. It is evident that sexual harassment encompasses both unwelcome comments and the non-consensual receipt of sexually explicit material. This includes pornographic material produced by someone else or nude images of the offender. The latter offence may additionally be construed as indecent exposure in domestic law. Sexual comments or the sending of sexually explicit material is not offensive *per se* but centres on the lack of consent of the recipient. Thus, the fact that communication is unwelcome is key and consent must be contextually assessed, where social and cultural factors may have an impact on the type of behaviour that is considered unwanted. The context of the Internet may, to an extent, affect such assessments.

[192] Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (2006), OJ L 204/23, para. 2 (d): 'where any form of unwanted verbal, non-verbal or physical conduct of a sexual nature occurs, with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment.' While linked to the workplace, as noted, the EU Parliament has called for the application of such to the Internet.

[193] CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 18.

[194] Art. 1 (a) of the International Labour Organization (ILO) Violence and Harassment Convention, 2019 (No. 190). *See* also UNGA, 'Intensification of Efforts to Prevent and Eliminate all Forms of Violence against Women and Girls: sexual harassment' (14 November 2018) UN Doc. A/C.3/73/L.21/Rev.1, para. 3: '[s]exual harassment encompasses a continuum of unacceptable and unwelcome behaviours and practices of a sexual nature that may include, but are not limited to, sexual suggestions or demands, requests for sexual favours and sexual, verbal or physical conduct or gestures, that are or might reasonably be perceived as offensive or humiliating'.

4.3.2.3 Online Sexual Harassment

Few definitions of "online sexual harassment" exist at the international level, given its novelty and that definitions of "sexual harassment" *per se* are rare in international law. Nevertheless, the EU has defined "cyber sexual harassment" as unwanted, offensive sexually explicit emails, photographs, text messages or advances on social networking sites or Internet chat rooms, or threats of sexual violence, with similar concepts employed in other soft law sources. [195] The definition of sexual harassment in the Istanbul Convention has also been affirmed as applying to cyber harassment. GREVIO has held that Art. 40 encompasses (1) non-consensual image or video sharing; (2) non-consensual taking, producing or procuring of intimate images or videos; (3) exploitation, coercion and threats (4) sexualised bullying; and (5) cyberflashing, while noting that this overlaps with sexist hate speech.[196] The UN Special Rapporteur on Violence against Women has also taken inspiration from this definition, discussing it in terms of '. . .any form of online unwanted verbal or non-verbal conduct of a sexual nature with the purpose or effect of violating the dignity of a person, in particular by creating an intimidating, hostile, degrading, humiliating or offensive environment'.[197] A distinction can be made between "direct sexual harassment" (for example, the non-consensual distribution of intimate photographs and rape threats against an individual), and "indirect sexual harassment" (spreading rape jokes and memes).[198] Sexist jokes are harmful in that the use of humour portrays demeaning views on women's unequal status as acceptable.[199] If involving hostility towards women as a group, it may also constitute sexist hate speech. With also the CEDAW Committee calling on states to take measures against online sexual harassment and the EU extending its approach to harassment to the Internet—aiming to harmonise the criminalisation of "cyber harassment" in Member States—steps are thus increasingly being taken to transpose the concept of sexual harassment to the online environment. However, certain elements in the current approach require particular examination in relation to the Internet.

---

[195] European Parliament, Directorate-General for Internal Policies of the Union, 'Cyber violence and Hate Speech Online against Women' (2018), p. 15. *See,* also, EIGE, 'Cyber Violence against Women and Girls' (2017), p. 2; OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 39.

[196] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', paras. 38–39.

[197] UNHRC, '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), para. 40. Meanwhile, Powell and Henry define "online sexual harassment" as '. . .offensive, humiliating or intimidating conduct that is unwanted or unwelcome and of a sexual nature', but does not aim to define it for legal purposes. *See* Powell and Henry (2017), p. 154.

[198] A distinction based on MacKinnon's delineation between "quid pro quo" and "hostile environment" harassment. *See* MacKinnon (1979). This may also be described as "ambient harassment". S*ee* Glomb et al. (1997).

[199] Bell (2021), p. 169.

Since sexual harassment regulation arises from and primarily has been applied at the domestic level, it may inform the transposition of the concept to the Internet, particularly from the viewpoint of liability. As mentioned above, states are obliged to adopt criminal legislation prohibiting acts of sexual violence, thus holding individual perpetrators accountable. This applies regardless of the sphere. Meanwhile, domestic sexual harassment laws extend liability to agents of control, which would be the equivalence of intermediaries on the Internet. However, the legal framework concerning sexual harassment at the domestic level is often narrowly construed in terms of the contexts and parties involved.[200] Not only must the activity transpire in this setting but domestic laws also tend to assume that the subsequent harm occurs in the same protected context. [201] That is, online harassment perpetrated by a stranger which leads to an individual withdrawing from the Internet and thus affects his/her employment opportunities would commonly not be subsumed.

As noted, while international sources indicate a broader approach to sexual harassment, some still limit this offence to certain environments. However, not only does this uphold a narrow approach to the areas of "work" and "education", it also undermines the importance of applying the sexual harassment rhetoric to other spheres with public characteristics. First of all, work and education are no longer limited to physical locations but may be conducted through e-commerce and online education. Internet users network online, post CVs and search for jobs and information relevant to their education. Sexual harassment in the context of the workplace increasingly occurs through digital communication, for instance, through emails or messages on electronic bulletin boards.[202]

Even in instances where the harassment is not in the form of a work email from an employer or colleague, there is room to argue in favour of a conception of sexual harassment that moves beyond the traditionally protected settings and that acknowledges that sexual harassment in one setting may produce harm in another.[203] Sexual harassment online may in effect be detrimental to the economic and social rights of women, since it may lead to either a retreat from or decreased use of the Internet. Furthermore, as noted by Mary Anne Franks, even if a person withdraws from the Internet, harassing comments about an individual may still be posted. The harm thus arises not solely through the victim's awareness of the post but also, for example, in tainting the reputation of a person.[204] This expansive view considers acts occurring outside a protected setting but that produces effects in that sphere, such as a withdrawal from the Internet, that is, not a direct spatial or relational connection.

Regardless of the implications for the working lives of victims, the Internet as an important public sphere in itself should be considered. As states incur obligations to ensure Internet access under international human rights law—in view of its

[200] *See* generally Roth (1999), p. 56.

[201] Franks (2012), p. 655.

[202] Harris et al. (2005), p. 74.

[203] Franks (2012), p. 656.

[204] ibid., p. 682.

importance to democracy and the fulfilment of a range of human rights—the Internet as a public sphere should be on a par with the more traditional settings of sexual harassment laws. In considering the rationale of sexual harassment laws as applying to settings closely connected to the goals of gender equality, characterised by a captive audience, this could certainly extend to the Internet. Captivity in this manner involves the importance for individuals in maintaining an Internet presence. As noted above, the European Parliament has called for the prohibition of sexual harassment in all public spaces, also denoting the Internet as such. Nevertheless, the particular characteristics of the Internet have also been raised as factors limiting the applicability of harassment laws to this sphere. Its benefits as a space for spirited debate, including harmful opinions, has in this regard been emphasised. This affirms the ideological view of cyberspace as distinct and subject to a broader freedom of expression than in the physical world.[205]

The narrow construction of spheres in domestic laws is also related to the issue of liability, which is limited to a narrow category of agents, being those deemed capable of exerting effective control over the setting.[206] Such laws do not generally involve liability for the harasser but for the employer. In the US, the employer may be held responsible not only for the behavior of employees but also third parties over which the employer has control, such as customers.[207] Transposing such an approach to the Internet and international human rights law presents certain challenges. As noted in Sect. 3.4, state obligations under international human rights law may, under narrow circumstances, include the regulation of Internet intermediaries. A benefit of applying the domestic sexual harassment framework to the Internet is a strengthening of the argument for developing state obligations to adopt laws extending liability to such corporations. For example, Mary Anne Franks proposes that a third-party liability regime, similar to traditional sexual harassment laws, should be placed on website operators, as an analogy to employers.[208] Companies may in fact be more responsive to control in cyberspace than in many instances of harassment in IRL, as evidence is recorded and clearly identified agents of control are provided on websites.[209] Although assessing awareness in relation to intermediary liability is contentious *per se*, principles of knowledge suitable to the context and the content may be developed.[210]

From the perspective of state obligations in international human rights law, technical features of the Internet must be borne in mind when considering the ability to control the online environment, informed by the type of website and the form of harassment. This affects the approach to both intermediary and individual liability.

[205] Volokh (1996), p. 414, at least involving "hostile environments".

[206] Franks (2012), p. 659.

[207] *Lockard v Pizza Hut, Inc.,* 162 F.3d 1062, 1073 (10th Cir. 1998).

[208] Franks (2012), p. 656.

[209] ibid., p. 683.

[210] ibid., p. 687. According to Franks, actual rather than constructive knowledge would be appropriate.

As mentioned previously, in view of the specific characteristics of the Internet, the right to privacy must include the notion of a portable private space, focusing on subject matter and steps taken to control privacy.[211] Since the spatial aspect of the right to privacy is diminishing, an autonomy-based approach is promoted in this context.[212] As such, it is argued that state regulation of unwanted, harmful messages is reasonable, whether it involves spam[213] or individual messages.[214] However, the level of control may differ depending on whether it concerns comments to a broader public—for example, when published on an open website—spam or individual private messages. Restrictions of one-to-one messages also have a more limited impact on the freedom of expression than restrictions of one-to-many communications, as the latter inhibits information also to willing listeners.

Although not affirmed as an obligation for states, it is implied that a notice-and-takedown mechanism may be a suitable form of intermediary liability *vis-à-vis* comments on open fora. As discussed previously, state obligations to regulate Internet intermediary and media publisher liability has in international human rights law and EU law primarily been confirmed in relation to hate speech and child pornography, in view of the level of harm and the ability to assess the legality of such forms of speech. EU law currently excludes harassment from the concept of "illegal" content, albeit generally categorised as "harmful", which indicates certain obligations of due diligence for Very Large Online Platforms (VLOPs) in the proposed Digital Services Act (DSA), but not general obligations of removal for intermediaries. Nevertheless, the proposed EU directive on violence against women places obligations on states to criminalise cyber harassment, albeit defined narrowly, which would extend intermediary liability to such conduct. Meanwhile, in the recent case of *Høiness v Norway* of the ECtHR, speech which could be categorised as sexual harassment was not deemed to generate additional obligations to regulate the liability of website operators beyond its notice-and-takedown system, as the speech was not "clearly unlawful".[215] Such an approach implies that sexual harassment is more subjectively determined, with its assessment beyond the capabilities of intermediaries or media publishers. However, whereas certain forms of communication may involve complex assessments for intermediaries, for example, on non-consent, it is possible to establish many forms of sexual harassment objectively. Additionally, contextual aspects are relevant in legal assessments in relation to a range of offences, including hate speech. As noted in the previous chapter, this includes cultural and social influences on harm, the particular website, such as its audience and jargon, as well as preceding messages. Nevertheless, mainly state obligations *vis-à-vis* the individual perpetrator in relation to these forms of speech arise in international human rights law.

[211] Slane (2005), p. 258.
[212] ibid., p. 267.
[213] ibid.
[214] Volokh (1996), p. 411.
[215] *Høiness v Norway* App no. 43624/14 (ECtHR 19 March 2019).

Although it is unlikely that spam qualifies as a form of sexual harassment even when it includes sexual content, the case of *Muscio v Italy* of the ECtHR is informative on the issue of receiving unwanted sexually explicit material on the Internet and the right to privacy.[216] The case involved the receipt of pornographic content in the form of spam. The domestic authorities discontinued proceedings initiated by the applicant as there was no indication of defamation, fraud or the unlawful use of the applicant's data, regulated under domestic law. Additionally, the material, although pornographic, was not obscene. The Court acknowledged that inferring state responsibility *vis-à-vis* unsolicited mail is possible in certain instances, given that IT network operators act in the framework of agreements with state authorities and under their supervision. Accordingly, if the applicant believed that negligence could be imputed to the state or the operator to whose service he subscribed, he could have brought an action for damages before civil courts on account of a lack of supervision and/or efficient protection against the dispatch of unsolicited e-mails.

While the ECtHR considered the receipt of unwanted or offensive communications an interference of a person's private life, it argued that once connected to the Internet, the private lives of users of electronic mail systems enjoy more limited protection, due to being exposed to unwanted messages, images and information. Additionally, it is in practice difficult for states to combat spam and trace senders. Solutions instead include such options as users installing filters. The case was thus declared manifestly ill-founded. Consequently, the case affirms the theoretical possibility of inferring state responsibility in cases of negligence in regulating the Internet environment but also technical difficulties in controlling one-to-many communication, such as spam. This affects the content of state obligations. Furthermore, the case implies that the scope of the right to privacy in certain regards is different on the Internet. Again, technical difficulties in controlling offensive messages lower "reasonable expectations" of privacy, that is, the environment has an impact on the boundaries of the private sphere as well as social norms on privacy. Thus, while the CEDAW Committee has defined sexual harassment as encompassing the unwanted display of pornography, it is unlikely that it would involve spam, for the same reasons.[217]

Offensive spam may be differentiated from interpersonal communications of images. Both men and women receive unsolicited pornographic material through advertising/spam. However, women are more often sent explicit images from other private individuals, which may also constitute a sexual offence.[218] Both forms of acts involve forcing someone to view pornography and are thus invasions of privacy.[219] Nevertheless, it is more feasible to regulate one-to-one messages, such as personal email, as they insulate the unwilling listener, than restricting one-to-

[216] *Muscio v Italy* App no 31358/03 (ECtHR, 13 November 2007).

[217] CEDAW, 'General Recommendation No. 19 on Violence Against Women', para. 18.

[218] McGlynn (2022), p. 3.

[219] McGraw (1995), p. 517.

many speech, including spam, as the assessment of its non-consensual receipt and delineations of liability may be unclear.[220] At the domestic level, harassment laws frequently require that the perpetrator intended to harass or humiliate the victim, which may similarly be applied to online incidents.[221] Naturally, this may be difficult to ascertain in cases of spam. Furthermore, the level of "reasonable expectations" of privacy appears to differ. In relation to interpersonal harassment, the most appropriate form of regulation is thus holding the harasser accountable, similar to the regulation of sexual violence. That is, obligations to investigate and prosecute the individual perpetrator arise. When interpersonal harassment occurs in private pockets of the Internet, the control of intermediaries is limited.

Complexity still ensues in how to assess whether messages are "unwanted" in the online environment.[222] Non-consent cannot solely be presumed on the basis of its content, for instance, nude photographs, in the context of such current social practices as "sexting". As mentioned in relation to the theoretical discussion on the right to privacy, certain subject matters, such as sexuality, are considered aspects of the right to private life. Thus, regulation may not categorise certain sexual practices as privacy invasions *per se*. As in relation to sexual violence in general, an assessment of the consent of the recipient is thus central. This may be expressed or implicit and may additionally consider whether the offender reasonably believed that the recipient consented.[223] Presumably, the identity of the sender is relevant, that is, photographs from a stranger will generally be assumed to be indecent exposure, while they would require a more complex assessment when sent, for instance, from a partner or acquaintance. In contrast to such consent-based delineations, a "motive-based" cyberflashing offence has been proposed in the United Kingdom, which focuses on whether the perpetrator intended to cause humiliation or to obtain sexual gratification, and was reckless as to whether the victim would be caused alarm, distress or humiliation.[224] As these issues have not been dealt with at the international level, it is primarily a matter for consideration in domestic law. Nevertheless, general standards in international human rights law *vis-à-vis* the assessment of consent in sexual interactions apply, such as the necessity of adopting a contextual and gender-sensitive approach.

[220] Volokh (1996), p. 421.

[221] *See* comparative law in Cobb (2020).

[222] Several options for domestic laws on cyberflashing are plausible, including (1) a consent-based standard; (2) a motive-based standard or; (3) a consent- and motive-based standard. *See* McGlynn (2022), p. 4.

[223] McGlynn (2022), p. 12.

[224] Law Commission of the UK, 'Modernising Communications Offences – A final report' (20 July 2021), para 6.133.

4.3.2.4 Conclusion

According to international human rights law, sexual harassment is considered a form of gender-based violence and sex discrimination. This corresponds to linguistic theories and the feminist discourse on gender equality. Its prohibition is both explicitly regulated and included in general provisions in international human rights law, such as the right to privacy, which offers protection of psychological integrity and sexual autonomy. Whereas the approach to sexual harassment at the domestic level commonly is contextually narrow and this, to a certain extent, is reflected at the international level, several sources in international human rights law take a broader approach, focusing on the nature, purpose and harm of acts or speech. From such a standpoint, provisions concerning sexual harassment apply also to the Internet.

The underlying rationale of ensuring gender equality in the public sphere, and elements such as a captive audience and means of exercising effective control, are certainly applicable to the online sphere, providing room for the extension of sexual harassment laws both domestically and internationally. The most challenging element is that of liability, related to the ability to control content. General human rights law provisions, such as the right to privacy, place obligations on states to regulate conduct, be it through civil or criminal law, depending on the severity of the act, and to hold individual perpetrators accountable. Meanwhile, the application of the domestic approach to sexual harassment provides an argument for extending liability also to agents of control, such as online platforms. Nevertheless, according to the ECtHR, certain forms of communication, such as spam, are difficult for states to control, thus mainly placing the responsibility on Internet users themselves to block content, for example, through filters. Similarly, interpersonal messages, such as through email or private messages, can sorely be prevented through technical means, beyond user restrictions. This entails that the regulation of the liability of perpetrators remains the main venue in terms of obligations to protect. As sexual violence in general places stricter obligations on states, requiring a criminalisation of the acts, as opposed to sexual harassment which may involve either criminal or civil law, the categorisation of online communication is thus relevant. Furthermore, whether liability arises, and in what form, *vis-à-vis* publications on public message boards depends on various factors, such as the nature of the speech and the type of website and platform. Requirements to regulate secondary liability, extending beyond a notice-and-takedown system and use of moderators, was rejected by the ECtHR in *Høiness v Norway* in relation to online harassment on public message boards, in contrast to cases on hate speech. As will be argued in Sect. 4.4, if certain forms of sexist speech were to be categorised as hate speech, enhanced levels of liability for website operators would thus ensue.

Complexities also arise in assessing what forms of speech constitute sexual harassment. The few definitions of sexual harassment at the international level mainly focus on unwanted interactions of a sexual nature, producing an environment harmful to the dignity and psychological or physical well-being of the individual. It is clear that it encompasses speech, such as comments and images. Although

non-consent is not an aspect in most domestic laws on sexual harassment, as it is superfluous in vertical relationships, it becomes relevant in the context of the Internet. A contextual assessment is necessary in this regard, as the online environment may have an impact on the evaluation of harm. As mentioned above, the ECtHR has indicated that "reasonable expectations" of privacy are reduced on the Internet, and the pervasive use of coarse language requires greater tolerance by users. Furthermore, new sexual practices through ICTs, such as sexting, must also be borne in mind. At the same time, care must be taken not to raise the threshold of acceptable conduct in this context as it would particularly disadvantage women, who are disproportionately affected. With the Internet denoted as, at a minimum, a quasi-public sphere, protection against such harm in this environment must be equivalent to standards applied in the physical world.

4.3.3 Threats of Violence and Disclosure

Threats of violence on the Internet include death threats and threats of sexual or other physical violence involving the recipient or family members. Threats may also concern the distribution of personal information or images. It occurs in a variety of situations. For example, online threats are commonly directed at women active in the public sphere, such as female human rights activists, politicians and journalists, in order to silence them or damage their credibility.[225] It may also transpire in the context of sexual harassment, stalking or domestic violence from persons known or unknown. Such threats are both published on public message boards and in the form of private messages. These forms of threats may cause overwhelming fear for the safety of the victim or family members, in addition to producing broader effects, such as impeding women's involvement in the public sphere.[226]

Depending on the circumstances of the threats, the approach varies in international human rights law. This includes the nature of the threat, the setting in which it occurs and the identity of the perpetrator. At a general level, both the IACmHR/IACtHR and the ECtHR have held that threats of harm involving acts contravening the prohibition on torture, inhuman or degrading treatment may constitute violations

[225] UNGA, 'Promotion of the Declaration on the Right and Responsibility of Individuals, Groups and Organs of Society to Promote and Protect Universally Recognized Human Rights and Fundamental Freedoms: Protecting Women Human rights Defenders' (30 January 2014) UN Doc. A/RES/68/181; UNHRC, '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), paras. 28–29.

[226] An IPSOS Mori poll on online harassment in eight countries, commissioned by Amnesty International in 2017, found that at least 41% of women who had been abused online feared for their physical safety, and 24% feared for their family's safety, since online mobs who attack women often issue detailed and graphic threats against their children. *See*, <https://drive.google.com/file/d/1-gxSWRJsEl-CCO4HGs4uqV6NNoYe_nP2/view> Accessed 8 March 2022.

of the same, given the psychological harm caused.[227] While the IACtHR has categorised threats as forms of psychological torture in several cases, the ECtHR has considered them as 'at least inhuman treatment' but, depending on 'the circumstances of a given case, including, notably, the severity of the pressure exerted and the intensity of the mental suffering caused', they may reach the threshold of torture.[228] For example, sexual violence falls within the scope of this provision and the prohibition thus encompasses rape threats. Nevertheless, threats of rape have solely been addressed in cases involving state actor perpetrators in detention settings and in conjunction with other acts, resulting in findings of torture.[229] As torture is defined as an act perpetrated by a state actor, for a particular purpose, this provision has in practice been excluded in the context of harm between private actors.[230] If the threats are less grave and/or involve private actors, they may in certain limited circumstances constitute inhuman or degrading treatment.[231] Nevertheless, threats between private individuals more commonly engage the right to privacy.

Different types of obligations ensue depending on the identity of the perpetrator, mainly involving positive obligations in relation to private actors. As a first step, this requires the criminalisation of the various forms of threats. A common issue in relation to online threats is that they are not criminalised—as opposed to consummated acts—or are trivialised where domestic laws are applicable. As noted above, threats involving acts other than violence are seldom prohibited, such as the threat of distributing intimate images.[232] Nevertheless, in order to effectively prevent crime, it is necessary to criminalise also threats, a standpoint affirmed by the ECtHR.[233] A positive obligation to take preventive operational measures to protect individuals also ensues.[234] In order for such an obligation to arise in cases of a threat made by a private actor, it must be demonstrated that the authorities knew or ought to have known of a 'real and immediate' risk of ill-treatment.[235] The state must consequently

[227] *Urrutia v Guatemala* (merits, reparations and costs) IACtHR Series C No. 103 (27 November 2003), para. 92; *Campbell and Cosans v the United Kingdom* (1982) 4 EHRR 29, para. 26; *Gäfgen v Germany* (2010) 52 EHRR 45, para. 91.

[228] *Gäfgen v Germany*, paras. 91, 108.

[229] *Case of Maria Elena Loayza-Tamayo v Peru* (IACtHR), para. 58; IACmHR, 'Report on the Situation of Human Rights in Haiti,' OEA/Ser.L/V/II.88 Doc. 10 rev (1995), para. 134: 'Rape and the threat of rape against women also qualifies as torture in that it represents a brutal expression of discrimination against them as women'; *Prosecutor v Miroslav Kvocka*, ICTY, Case No. IT-98-30/1-T, Judgment of 2 November 2001, para. 561 (although implicit).

[230] Art. 1 of the United Nations Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (1984) GA res. 39/46, UN Doc. A/39/51, 1465 UNTS 85, entered into force 26 June 1987.

[231] *Volodina v Russia* (ECtHR).

[232] UNHRC, '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), para. 82.

[233] *Beizaras and Levickas v Lithuania* App no 41288/15 (ECtHR, 14 January 2020), para. 111.

[234] *Osman v the United Kingdom*, para. 115.

[235] ibid., para. 116; *Kontrova v Slovakia* App. no. 7510/04 (ECtHR, 31 May 2007).

do everything that can reasonably be expected to prevent the fulfilment of the threat. Threats also trigger the obligation to conduct effective investigations, that is, a procedural aspect of the right in question.[236]

Threats of physical violence have been addressed by the ECtHR in cases of domestic violence, that is, interpersonal offences, generating positive obligations to prevent violations. This has been analysed as an aspect of Article 2 in cases resulting in the death of the victim,[237] Article 3 if in conjunction with physical violence[238] and Article 8 in other instances. In *Hajduova v Slovakia*, threats of violence were *per se* considered to constitute a violation of Article 8, viewed as causing mental anguish.[239] The Court held that the threats against the applicant were to be considered in light of the history of physical abuse and menacing behaviour within the family, with such threats arousing in the applicant a well-founded fear that they might be carried out.[240] This affected her psychological integrity and well-being. The question of whether the failure to react to death threats alone could be a violation of Article 2 arose in *A. v Croatia* but the Court chose to limit its review to Article 8, as the threats were not considered sufficiently severe.[241] The CEDAW Committee has similarly criticised states for failing to promptly, adequately and effectively investigate complaints involving death threats or threats of violence, in the context of domestic violence, from the viewpoint of gender discrimination.[242] Furthermore, the ECtHR has in several cases considered that ineffective investigations of harassment by private individuals, including verbal assault and physical threats motivated by racism, constitute violations of the right to privacy, noting the increasingly high standard involved in the protection of human rights.[243] In this regard, only efficient criminal law mechanisms can ensure adequate protection and serve as a deterrent factor.[244]

Moreover, the concept of "violence against women" includes threats of physical, sexual, psychological or economic harm[245] and the Istanbul Convention considers threats to be a form of psychological violence.[246] The CEDAW Committee has held that gender-based violence does not require a direct and immediate threat to the life

[236] *Volodina v Russia* (ECtHR).

[237] *Opuz v Turkey* (2010) 50 EHRR 28.

[238] *Volodina v Russia* (ECtHR).

[239] *Hajduova v Slovakia* (2011) 53 EHHR 8.

[240] ibid., para. 49.

[241] *A. v Croatia* (2015) 60 EHRR 26.

[242] *O.G. v. the Russian Federation*, CEDAW Communication No. 91/2015 (6 November 2017), para. 7.6.

[243] *R.B. v Hungary* App no 64602/12 (ECtHR, 12 April 2016), para. 84; *Alković v Montenegro* App no 66895/10 (ECtHR, 5 December 2017), para. 69.

[244] *Beizaras and Levickas v Lithuania* (ECtHR), para. 128.

[245] Art. 3 (a) of the Istanbul Convention; CEDAW, 'General Recommendation No. 19: Violence Against Women', para. 6; CEDAW, 'General Recommendation No. 35 on Gender-Based Violence against Women', para. 14.

[246] Art. 33 of the Istanbul Convention.

or health of the victim.[247] The UN Special Rapporteur on Violence against Women has specifically categorised online threats—including threats of physical and/or sexual violence—as gender-based violence.[248] Also the ECtHR considers threats a form of psychological abuse in the context of gender-based violence, and that 'vulnerable victims may experience fear regardless of the objective nature of such intimidating conduct'.[249] Accordingly, this is not restricted to contexts of domestic violence. As a form of violence against women, it engages broader obligations of prevention, such as taking measures to strengthen gender equality and combat gender stereotypes.

It should also be noted that threats against the physical integrity of individuals have been addressed in a similar fashion as hate speech in terms of allowing states to place liability on online news portals, that is, to restrict speech in order to protect the rights of others.[250] Both forms of speech have been categorised as "clearly unlawful".[251] Nevertheless, the ECtHR has not considered that threats against an individual *per se* involve an exclusion from the freedom of expression, through Article 17, unless involving language 'aimed at weakening or destroying the ideals and values of a democratic society'.[252] If the threat constitutes incitement to violence, that is, is directed at a protected group or an individual as part of such a group, it may be excluded from protection. For example, hateful comments on social media directed against particular individuals, such as members of the LGBTQ community, have been considered hate speech and/or incitement to violence.[253]

Finally, threats may involve acts other than physical violence, such as threats of the dissemination of private information, such as intimate images. As noted, this can be categorised as a form of image-based sexual abuse or sexual harassment. Given the approach by regional human rights law courts that threats of violence encompassed by such provisions also constitute violations of the same, it can be argued that since non-physical sexual violence and the disclosure of personal information contravene the right to privacy, threats of that nature may also be included. For example, in *Khadija Ismayilova v Azerbaijan*, the threat of the release of images of a sexual nature causing public humiliation discharged the positive obligation of the state to investigate the identity of the anonymous letter and whether it was connected to the applicant's professional life.[254] Meanwhile, the proposed EU

[247] *V.K. v. Bulgaria*, CEDAW Communication No. 20/2008, UN Doc. CEDAW/C/49/D/20/2008 (27 September 2011), para. 9.8.

[248] UNHRC, '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), para. 27.

[249] *Volodina v Russia* (ECtHR), para. 98.

[250] *Delfi v Estonia* (2014) 58 EHRR 29, para. 159.

[251] ibid., para. 153.

[252] *Rubins v Latvia* App no 79040/12 (ECtHR, 13 January 2015), para. 48.

[253] *Beizaras and Levickas v Lithuania* (ECtHR).

[254] *Khadija Ismayilova v Azerbaijan* (ECtHR), paras. 119–120.

directive on violence against women aims to harmonise the criminalisation of threats to produce or disseminate intimate images without consent among Member States.[255] The UN Special Rapporteur on Violence against Women has also called for the criminalisation of threats of the disclosure of intimate images, as a means of preventing the actual distribution of personal images.[256] Threats to distribute images may also be used as a form of coercion to compel an individual to perform a particular act, for example, as a means of sextortion, where the threat is subsumed by the subsequent offence. Threats of disclosure of the release of intimate images are at times also used as a means of coercion in domestic violence and human trafficking.[257]

Consequently, the identity of the perpetrator as well as the context have occasioned a distinction in international human rights law in terms of applicable norms. It is clear that contextual considerations in relation to interpersonal offences are made when assessing harm and whether a risk is "real and immediate", even though threats under certain circumstances may also cause psychological harm *per se*. It is likely that online threats of violence, be it sexual or other forms of violence, made by private individuals would be addressed as violations of the right to privacy in cases of state negligence to prevent or effectively investigate such incidents. At the same time, it is possible that threats on the Internet may not be considered equally severe, given the physical distance between the perpetrator and victim and in view of the pervasiveness of offensive language online. However, as mentioned previously, the fact that a threat is anonymous may also heighten the level of fear.

While not necessarily involving direct threats, expressions of fantasies of injuring women—be it women in general or a particular woman—also call into question legitimate restraints on speech. Sexual identity explorations have occasionally been conflated with action in domestic courts. For example, communicated fantasies of harm have been considered injurious to third parties, as speech has been construed as an initial step to physical harm.[258] Domestic laws in certain instances categorise such speech as an attempt or conspiracy to commit a criminal act and thus proof of criminal intent. It may also be considered a form of incitement to violence. Meanwhile, according to the commonly applied harm principle, actions should not be punished on moral grounds unless they cause actual harm.[259] This approach mainly focuses on tangible physical harm, not speculative harm nor offences to the sensibilities or moral commitments of others. Although there is a risk that thoughts or fantasies may result in harmful acts against a person, this is often speculative rather

[255] Art. 7 (c) of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[256] UNHRC, '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), para. 100.

[257] ibid., para. 32.

[258] Gilden (2016), p. 426.

[259] Mill (1859), p. 22.

than imminent. [260] From such a viewpoint, empirically supported justifications are necessary in order to criminalise expressions involving imagined harmful scenarios.[261] Otherwise the *ideas* expressed by speech rather than the actual effects are punished. Accordingly, in several domestic cases in the US, fantasies of harming women expressed online, where there was no indicated intent to carry out the threat, have been considered protected speech.[262] This again raises the question of whether intent to harm is viewed differently online. Since features of the Internet include the possibility of remaining anonymous and global communication takes place without regular social restraints, many users exaggerate their views and expressions, which may reduce the impact of statements. On the other hand, comments on the Internet may be considered particularly persuasive, given the impact of this forum on social norms. Since threats, thoughts or jokes on harming individual women or women as a group are common online, these are pertinent legal considerations.

While the latter form of speech has not been addressed in international human rights law, it is likely that the approach would depend on whether a description of harm involved an individual woman or women in general. The former, even if not constituting a direct threat, may be addressed as a form of gender-based violence that generates psychological harm, in contravention of the right to privacy. If the speech involves comments on harming women in general, it may similarly constitute gender-based violence and, plausibly, sexist hate speech in instances of incitement.

4.3.4 Defamation

4.3.4.1 Introduction

Defamation involves the dissemination of false statements. It is accordingly considered an invasion of the right to privacy, in the form of the "distortion" of the reputation of a person. As such, it may be a standalone offence or an aspect of online harassment. There are cultural differences among states concerning the extent to which they protect defamatory speech, related to varying theoretical approaches to the freedom of expression.[263] Autonomy-based speech theories protect expressions

[260] Gilden (2016), p. 458.

[261] ibid., p. 462.

[262] For example, the case of *United States v Valle* 807 F.3d 508, 513 (2d Cir. 2015) concerned a man who communicated his fantasies online to kill, dismember and eat women. The US Second Circuit Court concluded that the government may not punish individuals for their thoughts but solely their actions. It must thus refrain from criminalising '. . .an individual's expression of sexual fantasies, no matter how perverse or disturbing'.

[263] For example, in the US, defamation occurs solely if the defendant publishes information with 'actual malice', understood as knowledge that it was false, or with reckless indifference to their truth or falsity. In the UK, the defendant must prove the truth of allegations in material which is presumed false. *See* an overview in Koltay and Wragg (2020), part VI.

regardless of their veracity whereas, for example, theories emphasising democracy attach limited value to such speech, as it may mislead the public in its decision-making.[264] As will be seen, international human rights law allows but generally does not oblige states to regulate defamation. Broader variations in domestic approaches to the appropriate limits of speech are thus present in relation to defamation than *vis-à-vis* harm generating state obligations. However, in certain instances of grave harm to reputation, the protection of personal integrity may place obligations on states.

Defamation is not a gendered offence *per se* in terms of its nature. Nevertheless, various international organisations consider it a form of online gender-based violence.[265] This is in view of the content of defamatory speech involving female victims, which often centres on women's sexuality, such as promiscuity.[266] Women active in the public sphere, such as politicians, journalists, human rights defenders and women espousing feminist viewpoints, are also disproportionately targeted in comparison to men.[267] Meanwhile, more recently viewed in relation to the "Me too" movement, civil claims have been initiated against victims naming perpetrators of sexual violence and harassment. This has been noted by the UN Special Rapporteur on Violence against Women as a cause for concern.[268] It has been described by the UN Special Rapporteur on the Freedom of Expression as a form of 'weaponizing the justice system to silence women'.[269] It is also increasingly common for public figures to respond to harassment claims with defamatory statements concerning the victim, as a defensive strategy.[270] Domestic defamation laws may thus simultaneously provide remedies in cases of gender-based libel of women and be used to constrain female victims and activists.

---

[264] Oster (2017), p. 44.

[265] UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call' (23 October 2015), p. 22; OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 35; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 46.

[266] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 35.

[267] *See* generally UNHRC, '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), paras. 28–29; UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021) UN Doc. A/76/258, para. 21.

[268] UNHRC, '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), para. 31.

[269] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 22.

[270] UNHRC, '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), para. 31.

4.3.4.2 International Human Rights Law

The right to privacy protects such aspects of a person's life as honour, morals and psychological integrity.[271] The protection of reputation and honour is explicitly provided as an aspect of privacy provisions in several international treaties, including Article 17 of the International Covenant on Civil and Political Rights (ICCPR), which in Article 17 (2) places obligations on states.[272] Offensive speech may also be restricted in accordance with the legitimate aim of protecting the reputation of others in provisions on the freedom of expression, such as Article 10 (2) of the ECHR. In early cases before the ECtHR, the right to reputation was thus not considered a separate right but a legitimate aim among others permitting state interference.[273] However, in more recent cases, such as *Radio France v France*, the ECtHR has affirmed a right to reputation as an element of Article 8, as being part of personal identity and moral integrity in the context of social life.[274] Similarly, a person's honour is included.[275] An anomaly is thus created in protecting an individual's standing in *public* life within the right to privacy.[276] The right to reputation protects individuals against both defamation and the disclosure of personal information, the latter discussed in the following section.

Nonetheless, its recognition as an independent right has been sporadic and mainly affirmed in cases involving serious allegations, with an inevitable direct effect on the private life of the applicant.[277] As such, the right to reputation is only considered independent when an attack on a person's reputation reaches a certain level of gravity and causes such prejudice to the enjoyment of his/her right to private life so as to undermine his/her personal integrity.[278] For example, public accusations of violent criminal acts have been considered to damage an applicant's reputation of such a level of seriousness as to cause a prejudice to his right to privacy.[279] Also groups are protected from defamation, if involving negative stereotyping of a certain level of severity.[280] The Court has thus distinguished between the concepts of personal integrity—as inalienable and protected under human rights law—and

[271] *A v Norway* App no. 28070/06 (ECtHR, 9 April 2009), para. 73; *Ion Carstea v Romania* App no 20531/06 (ECtHR, 28 October 2014), para. 38. *See* also Art. 17 of the International Covenant on Civil and Political Rights (1966) GA Res 2200A(xxi), UN Doc. A/6316, 999 UNTS 171, entered into force 23 March 1976.

[272] *See* also Art. 11 (2) and Art. (3) of the American Convention on Human Rights (1969) O.A.S.T. S. 36, 1144 UNTS 123, entered into force 18 July 1978 (ACHR).

[273] *Lingens v Austria* (1986) 8 EHRR 407, para. 38.

[274] *Pfeifer v Austria* App no 12556/03 (ECtHR, 15 November 2007); *Einarsson v Iceland* App no 24703/15 (ECtHR, 7 November 2017).

[275] *Sanchez Cardenas v Norway* App no 12148/03 (ECtHR, 4 October 2007), para. 38.

[276] Barendt (2009), p. 59.

[277] *Karakó v Hungary* App no 39311/05 (ECtHR, 28 April 2009), para. 23.

[278] ibid., para. 23.

[279] *Einarsson v Iceland* (ECtHR), para. 52.

[280] *Aksu v Turkey* App nos 4149/04 and 41029/04 (ECtHR, 15 March 2012).

reputation, where the harm of speech depends on the social effects, addressed under defamation law.[281] Reputation *per se* does not engender as extensive protection. As a right, positive obligations ensue, for example, to provide victims of defamatory statements effective remedies.[282]

As an independent right, conflicts between the freedom of expression and the right to reputation often arise. Unlike the conflict between the freedom of expression and the protection against sexual violence within the right to privacy—which inevitably entails a balance in favour of the latter—harm to reputation is not considered as severe and thus creates a more complex equilibrium. There are inherent difficulties related to theoretical disagreements at the domestic level on the values of protecting privacy and the freedom of expression, often placing preeminent importance on the latter. In the jurisprudence of the ECtHR, the conflict is resolved through a fair balance test, involving an assessment of the scope and importance of each right in the context of the circumstances of the case.[283] From the standpoint of restrictions on the freedom of expression, the ECtHR in *Krasulya v Russia* explained in more detail that it would take the following into account when evaluating the effect on the reputation of the person: the position of the applicant; the position of the person against whom the defamatory statement was made; the subject matter of the publication; whether the allegations were facts or value judgments; the actual wording used by the applicant and the penalty imposed on him.[284]

As discussed previously, there is a hierarchy of protected speech which is apparent also in defamation cases. For example, the underlying aim of upholding democratic values generates a broad freedom of expression for politicians and the press. Both political speech and criticism of politicians thus receive extensive protection as the freedom of public debate is at the very core of a democratic society.[285] Politicians are expected to demonstrate a greater degree of tolerance to criticism than ordinary citizens, since they 'inevitably and knowingly' lay themselves open to close scrutiny.[286] A similar approach is taken in relation to public figures more broadly.[287] The acceptable level of public scrutiny and criticism is higher for public servants than for private individuals, but narrower than for

[281] *Karakó v Hungary* (ECtHR), paras. 22–23.

[282] *Standard Verlagsgesellschaft MBH v Austria (No. 3)* App no 39378/15 (ECtHR, 7 December 2021), para. 92.

[283] For example, *Von Hannover v Germany* (2005) 40 EHRR 1, para. 60; *Axel Springer AG v Germany* (2012) 55 EHRR 218, paras. 89–95.

[284] *Krasulya v Russia* (2007) 45 EHRR 40, para. 35.

[285] *Lingens v Austria* (ECtHR), para. 42.

[286] ibid., para. 42. *See* also UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)' (2011) UN Doc. CCPR/C/GC/34, para. 38.

[287] *Bodrožić v Serbia* App no 32550/05 (ECtHR, 23 June 2009), para. 54; *Von Hannover v Germany* (ECtHR), paras. 72–73; *Einarsson v Iceland* (ECtHR), para. 43.

politicians.[288] Meanwhile, the reputation of private individuals warrants the most extensive protection.[289] However, when private individuals involve themselves with issues of a public interest or concern, the Court has held that their privileged status is accordingly reduced.[290]

As the press fulfils an especially important role in democratic societies, in informing the public, it is afforded broader protection in defamation cases.[291] This extends to individuals or entities that do not belong to the press but perform a similar function, such as NGOs[292] or authors of books of public interest.[293] Given the value of the press, the extent of positive obligations to protect the reputation of an individual cannot be so wide as to have a chilling effect on the media.[294] Thus, when information is published in the media, the protection of individual integrity is balanced against the important function of, for example, media outlets, entailing that individual protection may be reduced if the matter is deemed to be of public interest.[295] A similar approach has been taken by the UN HRC.[296] This takes into consideration that people also enjoy the right to receive information, encompassed within the freedom of expression.[297] The "public interest" criterion has been interpreted to involve political issues, crimes, sports and performing arts.[298] A distinction is made between information that the public has a right to be informed of and publications intending to satisfy people's curiosity.[299] Even though public figures receive less protection of their privacy in this regard, personal matters such as marital or financial problems have been considered private.[300]

In defamation cases, the ECtHR furthermore considers whether statements involve facts or value judgments. The Court has drawn a distinction between expressions of "information" and "ideas or opinions", where the latter are considered value judgments that are impossible to prove.[301] The UN HRC has similarly held that defamation laws should include grounds for defence, such as the truth, and not

[288] For example, more extensive protection is provided to prosecutors, defence counsel and judges against defamatory speech than *vis-a-vis* other public servants, due to their roles as guarantors of justice. *See Nikula v Finland* (2004) 38 EHRR 45, para. 46.

[289] *Tammer v Estonia* (2003) 37 EHRR 43, para. 62.

[290] *Jerusalem v Austria* (2003) 37 EHRR 567, para. 39.

[291] *Lingens v Austria* (ECtHR), para. 41.

[292] *Vides Aizsardzības Klubs v Latvia* App no 57829/00 (ECtHR, 27 May 2004), para. 42.

[293] *Azevedo v Portugal* App no 20620/04 (ECtHR, 27 March 2008), para. 31.

[294] *Porubova v Russia* App no 8237/03 (ECtHR, 8 October 2009), para. 50.

[295] *Tammer v Estonia* (ECtHR), para. 62. *See* also *Pfeifer v Austria* (ECtHR).

[296] UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para. 47.

[297] *Lingens v Austria* (ECtHR), para. 41.

[298] *Von Hannover v Germany* (ECtHR), para. 60; *Somesan and Butiuc v Romania* App no 45543/04 (ECtHR, 19 February 2014), para. 25.

[299] *Von Hannover v Germany* (ECtHR), para. 65.

[300] *Standard Verlags GmbH v Austria* (2010) 50 EHRR 16.

[301] *Lingens v Austria* (ECtHR), para. 46.

encompass forms of expressions that are not, by their nature, subject to verification.[302] A defence of defamation is thus that the statement is true, since defamation solely involves false statements of facts. As stated in the *Lingens* case: '. . .the existence of facts can be demonstrated, whereas the truth of value judgments is not susceptible of proof'.[303] When categorising content, the ECtHR evaluates the context of speech in order to determine whether it involves statements of fact, or facts disguised as value judgments. Although it is primarily the task of national courts to make such assessments, the ECtHR may evaluate statements independently.[304] In cases where applicants have been unable to prove the veracity of factual statements in domestic courts, it has been deemed acceptable for the state in question to award damages to the impugned person.[305] At the same time, the Court has held that even if information is deemed to be true, statements may be unprotected if not considered relevant to the public interest, that is, a degree of protection against disclosure may still be required.[306] The assessment of protection against allegedly defamatory statements thus overlaps with cases on the disclosure of personal information.

In contrast, when involving value judgments, the speaker should not be required to prove that allegations are true. [307] However, in more recent judgments, the Court has held that value judgments still require an adequate factual basis, as opposed to precise allegations of facts in relation to factual statements. In *Petrina v Romania* and *Pfeifer v Austria*, the lack of proof of the factual allegation had a major impact on the finding of a violation.[308] However, the Court considered that even if the statements had been value judgments, the outcome would have been the same since the factual basis was insufficient to support the claim. The Court held that statements directly accusing a named individual, completely devoid of a factual basis, cannot benefit from the defence of exaggeration or provocation. Additionally, the Court has in other cases indicated that also value judgments have to be made in "good faith" in order to enjoy protection.[309] This does not mean that false statements are completely unprotected. The Court in *Salov v Ukraine* held that:

> . . .Article 10 of the Convention as such does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful. To suggest otherwise would deprive persons of the right to express their views and

[302] UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para. 47.

[303] *Lingens v Austria* (ECtHR), para. 46. Similarly, in *Steel and Morris v the United Kingdom* (2005) 41 EHRR 22, para. 93, the Court held that '[it is] not in principle incompatible with Article 10 to place on a defendant in libel proceedings the onus of proving to the civil standard the truth of defamatory statements'.

[304] *Rumyana Ivanova v Bulgaria* App no 3627/03 (ECtHR, 14 February 2008), para 57.

[305] *Harlanova v Latvia* App no 57313/00 (ECtHR, 3 April 2003).

[306] *Ion Carstea v Romania* (ECtHR), para. 34; *Somesan and Butiuc v Romania* (ECtHR), para. 26.

[307] *Lingens v Austria* (ECtHR), para. 46.

[308] *Petrina v Romania* App no 78060/01 (ECtHR, 14 October 2008), para. 50; *Pfeifer v Austria* (ECtHR), para. 48.

[309] *Castells v Spain* 14 EHRR 445, para. 46.

> opinions about statements made in the mass media and would thus place an unreasonable restriction on the freedom of expression set forth in Article 10 of the Convention.[310]

In this case, the statement of fact did not originate from the applicant. Rather, his statement consisted of personal assessments of factual information, also expressing doubt concerning its veracity.

The assessment of whether statements are value judgments or facts are complex and takes into consideration the preceding events, such as the applicant's conduct prior to the publication of the allegedly defamatory statement. For instance, in the case of *Egill Einarsson v Iceland*, the Court examined defamatory comments on an Instagram account, concerning a public figure.[311] Einarsson was a well-known person in Iceland, with controversial views on women and sexual autonomy, construed as encouraging sexual violence against women, opinions which he espoused also in public debates. In 2011 and 2012, two women reported to the police that they had been raped by Einarsson. The police investigations were discontinued as there was a lack of evidence. A leading newspaper published an interview with Einarsson in which he discussed the rape allegations, making disparaging comments about the alleged victims.

The same day, X published an altered photograph of Einarsson on Instagram, taken from the front page of the newspaper, with the caption "Fuck you rapist bastard".[312] In the photograph, X had drawn an upside-down cross on the applicant's forehead and written "loser" across his face. The Icelandic Supreme Court found in favour of X, considering the word "rapist" as an invective in the context of the case, involving a heated public debate, rather than a factual statement that the applicant had committed rape.[313] Meanwhile, the ECtHR found the statement objective and factual in nature, referring to a person who had committed the act of rape, in violation of domestic law. The veracity of the allegation could thus be proven. It did note that "rapist" may be considered a value judgment in certain situations. Whereas the domestic Supreme Court noted the context of the 'ruthless debate', according to the ECtHR, the domestic court had not considered the chronological link between the publication of the statement and the discontinuation of the criminal case on rape, as clearly the statement was linked to the criminal investigation.[314] Two judges issued separate opinions, noting that the context had changed the nature of the word "rapist" to become a value judgment—an invective—through the provocative statements of the applicant and/or alternatively the context of the statement, being the attached photograph.[315]

Furthermore, insulting speech is generally protected under Article 10, although the content, tone and form of statements must be considered. Here, a distinction is

[310] *Salov v Ukraine* (2007) 45 EHRR 51, para. 113.

[311] *Einarsson v Iceland* (ECtHR).

[312] ibid., para. 8.

[313] ibid., para. 14.

[314] ibid., para. 51.

[315] ibid., Dissenting opinions of Judge Lemmens and Mourou-Vikström.

made between value judgments—being allegations with some factual basis—and pure insults, considered abusive words with no individual meaning.[316] It has also held that wanton denigration falls outside the scope of protection.[317] Similarly, the ICCPR protects highly offensive speech, assessed in light of the context of statements.[318] The ECtHR has been especially accepting of defamatory statements made in oral exchanges or heated debate, since individuals are less likely to speak with careful consideration in such instances.[319] For example, the Court has held that speakers cannot reword, perfect or retract words during a demonstration, which affects the assessment.[320] This would entail that written statements online should be held to a higher standard. At the same time, as noted above, much online speech is reactive, similar to verbal exchanges. The potential impact of the damage is also considered in relation to the audience, for example, whether the statement was disseminated to the public or published in a forum with a limited readership.[321]

Whether defamation is regulated through civil or criminal law is relevant. The approach to the means of suppressing defamation is influenced by cultural and historical factors, similar to the scope of the freedom of expression in general. Commonly, states regulate defamation through civil law, whereas hate speech or threats of grave harm against an individual are often prescribed through criminal law. The criminalisation of defamation is thus disputed.[322] The ECtHR, the IACtHR, the UN HRC and the UN Special Rapporteur on the Freedom of Expression have indicated that it may be a disproportionate measure.[323] According to the IACtHR, such convictions chill free speech, especially when there is a public interest in exposing certain issues involving public officials, who must tolerate more intrusive

[316] For instance, in relation to the former, the Court has held that the use of certain terms such as "Nazi", does not automatically justify a restriction. *See Scharsach and News Verlagsgesellschaft mnH v Austria* 40 EHRR 569. However, the use of such words as "vampire" and "chief of gang of killers" have been deemed unacceptable. *See Lindon and Others v France* (2008) 46 EHRR 35, paras. 57, 66.

[317] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* App no 22947/13 (ECtHR, 2 February 2016), para. 76.

[318] *Malcolm Ross v Canada*, Communication No. 736/1997, UNHRC, UN Doc. CCPR/C/70/D/736/1997 (18 October 2000), para. 11.6.

[319] *Gavrilovici v Moldova* App No. 25464/05 (ECtHR, 15 December 2009), para. 58.

[320] *Birol v Turkey* App no 44104/98 (ECtHR, 1 March 2005), para. 30.

[321] *Pedersen and Baadsgaard v Denmark* App no 49017/99 (ECtHR, 17 December 2004), para. 79; *Raichinov v Bulgaria* App no 47579/99 (ECtHR, 20 April 2006), para. 48.

[322] From a US perspective, the effect of criminalising defamation '. . .is to chill speech. It does not promote the equality of persons or of ideas. It has no place in a democratic society.' *See*, for example, Lisby (2004), p. 435.

[323] In *Renaud v France* App no 13290/07 (ECtHR, 25 February 2010), para. 43, the criminal conviction of an individual for defamation was deemed disproportionate to the aim of protecting reputation. *See* also *Herrera-Ulloa v Costa Rica* (preliminary objections, merits, reparations and costs) IACtHR Series C No. 107 (2 July 2004). The case concerned a journalist in Costa Rica who was convicted of the crime of defamation after publishing allegations that a Costa Rican diplomat was involved in corruption.

publicity.[324] Nevertheless, criminalisation of 'the most serious cases' may be acceptable.[325]

4.3.4.3 Defamation on the Internet

Defamation is more likely to occur on the Internet than in other media due to user anonymity, distinct social norms online, possibilities of broadly distributing material and limited monitoring.[326] The Internet as the site for defamation in turn affects several aspects of its legal assessment. This includes the categorisation of public/private figures, the media and publishers. It also has an impact on the evaluation of harm to reputation. The delineation of public/private persons in view of their Internet presence has not been specifically addressed in international human rights law, where the status of individuals as public or private is fluid. For example, private individuals frequently—of their own accord—reveal information to the public through social media and may become well known through their media presence. Nevertheless, the ECtHR has in case law on hate speech considered whether persons were well-known bloggers or users of social media, from the perspective of the speaker. This indicates that the level of their online presence and following most likely will affect their categorisation.[327] In terms of the delineation of which fora are considered the media or who operates as a journalist, as noted, the Court has affirmed that blogs and influential social media users may fulfil the role of a "public watchdog", also noted in soft law sources in international human rights law.[328] Additionally, beyond news websites acting as publishers, the Court has taken a more extensive approach, evident in *Sanchez v France*, where a politician was considered a publisher through having set up a Facebook account open to the public.[329]

Domestic defamation laws generally allow for civil claims against the author of the comments or the publisher. These have also been affirmed by the ECtHR as the main venues of regulating liability for defamation, rather than intermediary liability.[330] Nevertheless, the Internet generates complexities in terms of determining who has authored content and the scope of secondary liability for online media

[324] *Herrera-Ulloa v Costa Rica* (IACtHR), para. 128.

[325] UN HRC, 'General Comment No. 34: Freedoms of Opinion and Expression (Art. 19)', para. 47; Davidson (2009), p. 207.

[326] Murray (2016), p. 217.

[327] *Savva Terentyev v Russia* App no 10692/09 (ECtHR, 28 August 2018), para. 81. However, this concerned the assessment of the impact of statements by persons accused of disseminating hate speech.

[328] *Magyar Helsinki Bizottság v Hungary* App no 18030/11 (ECtHR, 8 November 2016), para. 168; UN HRC, 'General Comment No. 34 on Article 19: Freedoms of Opinion and Expression', para. 44.

[329] *Sanchez v France* App no 45581/15 (ECtHR, 2 September 2021).

[330] Author: *Delfi v Estonia* (ECtHR), para. 147; publisher: *Orban and Others v France* App no 20985/05 (ECtHR, 15 April 2009), para. 47.

publishers. The mass media has traditionally employed a "one-to-many" form of communication, that is, it originates from a single source and is distributed to multiple recipients, where the publisher has editorial control over content. Meanwhile, social media is rather characterised by "many-to-many" communication. Various activities occur, such as posting, linking, re-posting and liking. The question thus arises whether a different approach to liability must be taken *vis-à-vis* online speech, for example, whether it includes the re-distribution of harmful expressions, in addition to the liability of the original author.

The ECtHR has touched on the issue of posting hyperlinks and "liking" content on the Internet in relation to defamation, where the Court has considered not only how such activities function, but also how regulation may affect the operation of the Internet. Hyperlinks are indications of the existence of another website and do not provide content of their own. In *Magyar Jeti Zrt v Hungary*, the ECtHR held that imposing objective liability for the posting of a hyperlink leading to defamatory content was a violation of Article 10.[331] In the view of the Court, hyperlinks could not be equated with dissemination and such a liability standard might have a chilling effect on the freedom of expression. It took into account that hyperlinks have a navigational function and are as such referencing tools, in that they do not communicate content. Furthermore, hyperlink providers do not create the content of the website nor control it, as it may be subject to change.[332] This does not entail that liability is excluded, but rather requires an assessment in the particular case, considering whether the person endorsed the content, repeated it or merely attached a hyperlink. Furthermore, whether the person could reasonably have known that the content was defamatory or otherwise unlawful must be borne in mind.

This is similar to the concept of a "single-publication rule" prevalent in many domestic laws, which entails that only the first publication of an article posted on the Internet may give rise to a claim of defamation and not subsequent downloads or repostings by users.[333] The risk with such a rule is that victims of defamation receive more limited protection. Nevertheless, the ECtHR in *Times Newspapers Ltd (Nos. 1 and 2) v the United Kingdom* held that a single publication rule is not a requirement of the Convention.[334] That is, a multiple publication rule is acceptable in defamation cases, meaning that each download or access to material is a new cause for a libel proceeding, involving liability also for redistributors. A similar finding has been made in relation to the disclosure of personal information, for the purpose of protecting individual integrity.[335] While such a rule strengthens access to remedies

[331] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 89. As the case involved a journalist, journalistic ethics were also considered.

[332] ibid., paras. 73–75. The viewpoint was based on the argument of dissenting judges in *Mouvement Raëlien Suisse v Switzerland* App no 16354/06 (ECtHR, 13 January 2011), Dissenting opinion of Judges Sajó, Lazarova Trajkovska and Vucinic.

[333] Smith (2007), para. 4-006.

[334] *Times Newspapers Ltd (Nos. 1 and 2) v the United Kingdom* App nos 3002/03 and 23676/03 (ECtHR, 10 March 2009), para. 49. This was also implied in *Lindon and Others v France* (ECtHR).

[335] *Aleksey Ovchinnikov v Russia* (ECtHR).

for victims of defamation, it may simultaneously have a chilling effect on the freedom of expression online.[336] Again, the conflict between rights *per se* is, in certain respects, heightened by Internet design.

In *Melike v Turkey* of the ECtHR, the status of "liking" content on the Internet was addressed in relation to the freedom of expression.[337] The applicant in the case had been dismissed from her employment at the Ministry of National Education subsequent to "liking" several posts by third parties on Facebook. The virulent posts alleged repressive practices by governmental authorities, including abuse of children in the school system, and criticism of certain political leaders and parties. These subject matters were considered political speech and of public interest, with limited room for restrictions by the state. Furthermore, the Court held that "liking" posts was a common form of exercising the freedom of expression online. While generally noting the potential global dissemination of statements online, it held that the impact of posts on websites or posts with a small readership is more limited. In assessing the effect of "liking" posts, it was not deemed equivalent to sharing content on social media, as it merely allows users to demonstrate their interest and approval of content, and it does not necessarily arise from an active wish to disseminate it. In this case, the posts had only attracted approximately a dozen "likes" and a few comments, and had thus not reached a large audience. The Court also considered the fact that the applicant was not renowned nor had the national authorities established whether pupils, parents and other employees had access to her Facebook account or had been able to view the "likes" through other parameters. In view of these factors, the dismissal violated her freedom of expression. However, arguably, "liking" may have broader effects than indicated in the case. For example, by supporting the credibility of information, affecting the algorithmic dissemination of material or by promoting similar content, it may influence the views of others.[338]

Furthermore, as discussed in Chap. 3, the ECtHR has in a line of cases on intermediary liability involving online defamation on social media and blogs held that libel proceedings against the authors of comments are the primary venues for accountability. At the same time, the practical difficulties in pursuing claims against individual perpetrators has been noted, given the prevalence of user anonymity. The ECtHR has even upheld the right for online newspapers to withhold data identifying anonymous contributors posting defamatory comments, albeit in the context of political debate.[339] Intermediary or media publisher liability is thus generally a more accessible route for accountability. However, such online platforms do not incur liability for third-party content to the same extent as traditional media

[336] CoE, Rapporteur Prévost E, 'Study on forms of liability and jurisdictional issues in the application of civil and administrative defamation laws in Council of Europe member states', DGI(2019)04, p. 6.

[337] *Melike v Turkey* App no 35786/19 (ECtHR, 15 June 2021).

[338] Mena et al. (2020).

[339] *Standard Verlagsgesellschaft MBH v Austria (No. 3)* App no 39378/15 (ECtHR, 7 December 2021).

publishers. Although online publishers are generally considered to be in greater control of content than intermediaries, including social media companies and blogs, these have been addressed in a comparable manner by the ECtHR in case law on defamation. The Court has in this regard mentioned defamation as an example of "unlawful" speech, alongside hate speech, to be removed by publishers.[340] However, whereas news websites incur obligations *vis-a-vis* hate speech, liability regimes for defamation have been deemed to largely fall within the scope of the margin of appreciation of states. Notice-and-takedown systems have in this regard been considered appropriate and in line with the general approach in international law. [341] Given the difficulty for AI to detect defamation and for human moderators to assess the veracity and legality of such content—especially in view of the lack of harmonisation at the regional and international level—particular forms of notice-and-takedown regimes have been adopted in certain jurisdictions.[342] This includes notice-wait-and-takedown systems, allowing time for counter notices.

In terms of harm, defamation is often exacerbated on the Internet. This includes the commonly coarse tone of communication as well as the longevity and global reach of comments. However, certain characteristics of the online environment arguably also reduce the impact of defamatory statements. As mentioned previously, it has been argued that the social acceptance of distinct norms on the Internet—with users more frequently viewing comments on the Internet with skepticism, given the prevalence of false information online—diminishes the harm to reputation, in view of its connection to social judgment.[343] Additionally, according to the UN Special Rapporteur on the Freedom of Expression, given the ability of individuals to exercise their right of reply instantly to restore the harm caused, the types of sanctions that are applied to offline defamation may be unnecessary or disproportionate.[344] In most cases it is possible for victims to respond to defamatory statements online, as opposed to in other contexts, such as newspapers and books. However, reputation is not necessarily restored through a reply, since the defamed person may not be aware of the derogatory statements. Furthermore, the utility of this measure is limited as the results on search engines are not necessarily chronological.[345] For example, on Google, material is permanent unless removed, and often without chronological progression of the debate. That is, even with rebuttals, the offending material may be the first item displayed in a search.[346] Moreover, a person

[340] *Delfi v Estonia* (ECtHR), para. 110.

[341] *Payam Tamiz v the United Kingdom* App no 3877/14 (ECtHR, 19 September 2017), paras. 54–56, 84.

[342] *See* in Angelopoulos and Smet (2016), p. 267.

[343] Nussbaum (2010), p. 104. *See* also *Payam Tamiz v the United Kingdom* (ECtHR), para. 81; *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 77.

[344] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (16 May 2011) UN Doc A/HRC/17/27, para. 27.

[345] Heller (2007), p. 281.

[346] ibid., p. 281.

may not wish to refute defamatory statements through a public reply. It is not uncommon that discussion boards or websites create a climate of fear that chill victims' willingness to rebut.[347] Even if a person replies, there is no guarantee that the message will reach the target audience or be believed.[348]

In case law of the ECtHR on defamation and intermediary liability, the impact of the Internet has in relation to the issue of harm mainly been addressed in the evaluation of the content, form and consequences of the impugned publication. In *Payam Tamiz v the United Kingdom*, the applicant argued that the limited possibility of holding an Internet intermediary liable in the domestic court violated his right to reputation in Article 8.[349] An entry on a blog called "London Muslim" published a photograph of Mr. Tamiz—a council candidate—and claimed that he had called girls in his hometown "sluts" in an old Facebook post.[350] Eight of the comments made were considered defamatory by the applicant. These were anonymous and included allegations that Mr. Tamiz was a drug dealer, promiscuous, racist and living in the UK on the basis of a false asylum claim.

The Court began by addressing the balance between Article 8 and Article 10. It affirmed that Article 8 encompasses the protection of a person's reputation, engaging both negative and positive obligations. However, it emphasised that positive obligations involving the sphere of private individuals generate a broad margin of appreciation for states, although with due regard of such factors as the gravity of the interference for the individual, the consensus among Member States and the fair balance struck by domestic courts between Article 8 and Article 10.[351] The latter point in particular engages a wide margin of appreciation for states when the balancing by national authorities has been conducted in conformity with the criteria established in the case law of the Court, requiring '. . .strong reasons to substitute its view for that of the national courts'. [352]

The Court reiterated that attacks on a person's reputation must attain a certain level of seriousness and must have been carried out in a manner causing prejudice to the personal enjoyment of the right to respect for private life. Notably, the Court observed that millions of Internet users post comments online each day and many express themselves in a manner that might be regarded as offensive or defamatory. However, 'the majority of comments are likely to be too trivial in character, and/or the extent of their publication is likely to be too limited, for them to cause any significant damage to another person's reputation'.[353] In this particular case, the Court agreed with the domestic courts that while the statements were offensive, for the large part, they constituted "vulgar abuse" of a kind '. . .which is common in

[347] ibid., p. 282.
[348] Rowbottom (2012), p. 376.
[349] *Payam Tamiz v the United Kingdom* (ECtHR).
[350] ibid., para. 7.
[351] ibid., paras. 78–79.
[352] ibid., para. 79.
[353] ibid., para. 80.

communication on many Internet portals'.[354] As a budding politician, he would be expected to tolerate such language.[355] Additionally, the Court argued that although the comments contained serious allegations, these would—in the context in which they were written—'likely be understood by readers as conjecture which should not be taken seriously'.[356] This in effect affirms the notion that a different set of norms operate on the Internet. The "Wild West" attributes of the Internet thus entail that individuals must accept vulgar statements to a higher degree than in other contexts, that is, this setting arguably lessens the degree of harm to the victim. It appears that the Court also categorised the applicant—a budding politician—as a public figure, thus further reducing the protection of his reputation.

Similarly, in *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* concerning a vulgar comment online, the Court argued that the '. . .offence may fall outside the protection of freedom of expression if it amounts to wanton denigration, for example where the sole intent of the offensive statement is to insult. . . but the use of vulgar phrases in itself is not decisive in the assessment of an offensive expression'. [357] Considering this in the context of the Internet, the Court held that '. . . the expressions used in the comments, albeit belonging to a low register of style, are common in communication on many Internet portals – a consideration that reduces the impact that can be attributed to those expressions'.[358] This also affirms the existence of distinct social norms on the Internet.

The context of the Internet was similarly noted by the Court in *Delfi v Estonia*, involving intermediary liability for comments constituting hate speech on a news portal.[359] The Court noted that the Internet provides an unprecedented platform for the exercise of the freedom of expression. However, '[d]efamatory and other types of clearly unlawful speech, including hate speech and speech inciting violence, can be disseminated like never before, worldwide, in a matter of seconds, and sometimes remain persistently available on line'.[360] Accordingly, '. . .while the Court acknowledges that important benefits can be derived from the Internet in the exercise of freedom of expression, it is also mindful that liability for defamatory or other types of unlawful speech must, in principle, be retained and constitute an effective remedy for violations of personality rights'.[361] As such, although the case in question involved hate speech, the Court addressed the matter generally in terms of an *obligation* to protect individuals from harmful speech, including defamation, rather than solely assessing the legitimacy of state interference. This entails that whereas

[354] ibid., para. 81.
[355] ibid., para. 81.
[356] ibid., para. 81.
[357] *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 76.
[358] ibid., para. 77.
[359] *Delfi v Estonia* (ECtHR).
[360] ibid., para. 110.
[361] ibid., para. 110.

the threshold of harm may be raised on the Internet, protection against defamation is still required in this sphere.

The extent of the readership also affects the assessment of harm, that is, whether published on a public or private website. In the above-mentioned case of *Egill Einarsson v Iceland*, the defamatory statements were published on an Instagram account.[362] X, according to his own statements, believed that only his friends and acquaintances, that is, his followers, would be able to view his posts. However, he had not set his settings to "private" and his posts were thus available to the public, which the Court addressed in relation to the consequences of the publication. It reiterated the benefits but also potential risks of the Internet to privacy, through its capacity to allow communication on a vast scale. Whereas defamation may occur also in situations where a post is available solely to a limited group of acquaintances, the fact that it was accessible to the public affected the assessment of harm. The fact that a Facebook account was set to "public" was also a significant factor in *Sanchez v France*, albeit concerning hate speech.[363]

4.3.4.4 Conclusion

Gender has an impact on defamation, involving a higher risk of harassment for women who contravene gender stereotypical norms of behaviour, such as being in positions of power or influence, including feminist activists, journalists and politicians. The fear of defamation may, as a consequence, impede women's involvement on political and social issues that entail visibility or authority. This undermines democracy and a gender equal public sphere. Furthermore, a majority of corporations search for information on applicants online during the hiring process and defamatory statements may influence their decision-making, thus also affecting the employment prospects of individuals. As mentioned above, empirical studies indicate that the type of defamation women are subjected to is also gendered, often involving aspects of their sexual life. Defamation has accordingly been categorised as a form of gendered online harm in soft law sources in international human rights law. Although defamation primarily has been addressed in cases evaluating the legitimacy of state interferences, the inclusion of protection of reputation in the right to privacy entails that positive obligations ensue in cases of substantial harm to the integrity of the individual. The ECtHR in *Payam Tamiz v the United Kingdom* and *Delfi v Estonia* also discussed positive obligations for states to protect individuals against reputational harm, in terms of access to remedies. This is a welcome development, given that defamation may generate grave consequences for a person's social life and his or her political and economic rights.

At the same time, legal proceedings have been brought against feminist activists and human rights defenders naming perpetrators of, for instance, sexual violence and

[362] *Einarsson v Iceland* (ECtHR).

[363] *Sanchez v France* App no 45581/15 (ECtHR, 2 September 2021).

harassment. Reactive speech—for example, evidenced in *Einarsson v Iceland*—that does not constitute mere value judgments, thus risks being subject to litigation. While allowing victims of gender-based violence to express their experiences may aid in shifting the discourse on harmful gender stereotypes, the use of domestic justice systems must remain the primary recourse for victims of harassment in terms of individual liability. Nevertheless, although international human rights law strikes a reasonable balance between privacy interests and the freedom of expression, domestic rules on intent, the burden of proof and the delineation of facts/value judgments vary greatly. It is thus important that such are not interpreted in a gendered manner, for example, in relation to elements such as "sufficient factual basis".

The approach to regulation of defamation in international human rights law is largely similar online/offline, involving consideration of factors such as the identity of the speaker and the victim, the nature of the comments and the context in which they were made. Nevertheless, the Internet affects the assessment of concepts, liability and the effects of speech. This includes the categorisation of public/private figures, authors and the media. Private individuals incur the most extensive protection, particularly in instances where other private individuals post defamatory statements, for instance, on social media. In the cases reviewed, defamatory statements were in the form of comments on a blog (*Payam Tamiz v the United Kingdom*), comment sections on news portals (*MTE and Index v Hungary* and *Delfi v Estonia*) and an Instagram post (*Einarsson v Iceland*), that is, not involving the press in the traditional sense. Categorising, for example, blogs as the media would entail that the interest of the public in accessing information would be a relevant consideration. Furthermore, all of the cases involved public figures, reducing the level of protection of their reputation. Although a distinction is natural, in view of the willingness to open oneself up to public scrutiny, it must also be noted that women more frequently than men raise fear of public attacks on their dignity as a factor deterring them from entering politics.[364] Since female politicians experience harassment to a greater extent, the reduced protection may in effect have an impact on the political involvement of women. Additionally, the acts of sharing, "liking" and hyperlinking content fuels the dissemination of harmful comments. Whereas several of these acts have been considered forms of expressions, the impact has not been deemed comparable to disseminating an opinion. However, this may be the case in certain contexts, for example, when hyperlinks involve endorsements or when "liking" posts on particular websites generates similar effects as distributing content.

Furthermore, as discussed previously, the ECtHR has in several cases been confronted with the issue of media publisher and intermediary liability in relation to third-party defamatory comments. A distinction has been made between online media outlets and traditional publishers, in view of the lower level of control of the former in relation to third-party content. Nevertheless, whereas the suitability of

[364] Bardall (2017), p. 107.

notice-and-takedown mechanisms has been affirmed in relation to both online media companies and intermediaries, such measures have not been construed as obligations, apart from an *obiter dictum* in *Delfi v Estonia*, despite the categorisation of defamation as "unlawful" speech. Such regulations currently fall within the margin of appreciation of states. Given the difficulty of intermediaries to assess the veracity of statements, such mechanisms are suitable. However, in view of the limited practical possibilities of holding individuals accountable, the ambivalence as to the categorisation of defamation as "clearly unlawful" and the potential positive obligations in terms of intermediary liability warrant clarification.

Several aspects of the assessment involve considerations of context, related to the concept of harm. The ECtHR in *Payam Tamiz v the United Kingdom* and *MTE and Index v Hungary* indicated that since the Internet spawns harsher language, reasonable expectations of protecting the reputation of individuals are reduced. That is, the mode of communication influences the tone of expressions, leading to more uninhibited, ill-thought-out comments and an informal style of writing. In the view of the Court, this entails that it is less likely that statements will be construed as true. This can be contrasted with the viewpoint of certain domestic courts, that user anonymity on the Internet enhances the risk that statements will be understood as true.[365] The Internet accordingly raises the threshold for acceptable speech, with speech more likely to be considered to cause offense than harm. As argued previously, vulnerable groups frequently bear the costs of unregulated harmful speech.

Furthermore, the Court in *Payam Tamiz v the United Kingdom* considered the size of the audience when assessing the harm to reputation, bearing in mind the actual readership of the website in question. This was similarly addressed at a general level in *Einarsson v Iceland*. Relevant factors include whether defamatory statements were published on an online newspaper, blog or social media, the privacy settings used, and whether the websites were well-visited. The ease with which such information may spread on the Internet has thus not been taken into account in relation to defamation. In contrast, in a case on hate speech, the ECtHR held that posts on social media are not *per se* less harmful than those on Internet news portals, given the risk of such "going viral".[366] This, to a greater extent, considers common communication flows on the Internet.

4.3.5 The Disclosure of Private Information

4.3.5.1 Introduction

The disclosure of personal information online is distinguished from defamation in that it involves information that is accurate. It includes private photographs and

[365] *See*, for example, *Vaquero Energy v Weir* (2004) ABQB 68 (Canada).

[366] *Beizaras and Levickas v Lithuania* (ECtHR), para. 127.

information, such as a person's address, health status and sexual orientation, with or without the intent of causing humiliation. This accordingly encompasses "doxing", which refers to the publication on the Internet of personal data with malicious intent, for instance, contact details. It also includes certain forms of image-based sexual abuse, such as "revenge pornography".[367] This may occur in the context of sexual harassment or sextortion and thus overlaps with previous discussions. In certain instances, it also falls within the scope of the broader concept of "cyber harassment", for example, as defined by the EU. Nevertheless, in the proposed EU directive on violence against women, this is confined to the publication by several individuals of "insulting material" through ICTs.[368]

Similar to defamation, the disclosure of private information is not gender-based *per se*. However, certain forms of disclosure are gendered, with harm to women's reputation mainly connected to sexuality and social norms of modesty. As argued above, the publication of intimate images may be categorised as a form of sexual violence, which is considered gender-based. Doxing and image-based sexual abuse have also been categorised as forms of violence against women online *per se*.[369] Statistics indicate that women are more likely than men to be victims of the non-consensual distribution of intimate photographs and information related to sexuality, such as sexual orientation and reproductive health data.[370] Several cases of doxing at the domestic level involve the disclosure of contact information alongside requests for sexual connections through, for example, fake advertisements, placing victims at risk of sexual assault.[371] Furthermore, similar to defamation, women in positions of authority or visibility as well as women's rights activists are targeted more frequently.[372] For example, the disclosure of personal information is one of the most common forms of online harassment that women in journalism are

[367] UNHRC, '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), para. 36; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', p. 30.

[368] Art. 9 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[369] EIGE, 'Cyber Violence against Women and Girls' (2017), p. 2; OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 33; UNHRC, '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), para. 33; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 38.

[370] Powell et al. (2020). For an overview, *see* Sparks (2021).

[371] LA Times, 'Former boyfriend used Craigslist to arrange woman's rape, police say' (11 Jan. 2010) <https://www.latimes.com/archives/la-xpm-2010-jan-11-la-na-rape-craigslist11-2010jan11-story.html> Accessed 17 March 2022. *See,* also, Citron (2014), p. 6.

[372] *See*, generally, UNHRC, '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), paras. 28–29.

subjected to.[373] Meanwhile, the non-consensual online dissemination of personal information, such as intimate images, is rarely explicitly criminalised at the domestic level. When regulated, the crime is often narrowly defined.[374] The following sections provide a general overview of the approach to disclosure of personal information. It is thus also applicable to disclosure of intimate images, although its categorisation as sexual violence entails additional considerations, addressed above.

4.3.5.2 International Human Rights Law

The right to protection of personal data, involving restrictions on gathering or publishing information without the consent of the individual concerned, is an aspect of the right to privacy in international human rights law, such as in the ECHR and the ICCPR. [375] This right is also included in the EU Charter of Fundamental Rights and has been affirmed in the Digital Rights judgment and the e-Commerce Directive of the EU.[376] Such EU-related instruments have in fact been mentioned in several cases of the ECtHR in relation to state interference through the collection and publication of personal data.[377] In this regard, principally state duties to respect privacy have been emphasised, that is, obligations to refrain from interfering through the collection of personal information, and positive obligations to protect individuals from the collection, processing and publication of such information by Internet Service Providers (ISPs). In contrast, little focus has been placed on the non-consensual publication of private material by other individuals through new technologies. This has mainly been summarily mentioned in reports on gender-based online violence.[378]

---

[373] UNESCO, 'World Trends in Freedom of Expression and Media Development: 2017/2018 Global Report', p. 156.

[374] UNHRC, '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), para. 82.

[375] Art. 17 of the ICCPR; UN HRC, 'General Comment No. 16: Article 17 (Right to Privacy), The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation' (8 April 1988) UN Doc. HRI/GEN/1/Rev.9 (Vol. I), para. 10; *M.L. and W.W. v Germany* App nos. 60798/10 and 65599/10 (ECtHR, 28 June 2018), para. 87. This has also been affirmed in UNHRC, 'Resolution 42/15 on the right to privacy in the digital age', UN Doc. A/HRC/RES/42/15 (7 October 2019). *See* also CoE Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Dara (ETS 108) 28 January 1981.

[376] Art. 8 of the Charter of Fundamental Rights of the European Union (the 'Charter'), 2010 O.J. (C83) 389; C-293/12 and C-594-12 *Digital Rights Ireland Ltd v Ireland* (2014) ECLI:EU:C:2014:238; 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 L 178/1.

[377] *Benedik v Slovenia* App no 62357/14 (ECtHR, 24 April 2018), para. 46–62; *K.U. v Finland* (ECtHR), para. 30.

[378] UNHRC, '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'

Protection against the disclosure of private information, similar to defamation cases, involves the safeguarding of personal identity and reputation, thus the social dimensions of the right to privacy. Through the jurisprudence of the ECtHR, it has been affirmed that private life covers '. . .personal information which individuals can legitimately expect should not be published without their consent'.[379] The ECtHR has emphasised that although truth is a defence against defamation, individuals may still be protected against the distribution of such information if harmful to a person's reputation and identity. The non-consensual publication of private information therefore creates a conflict between the freedom of expression and the right to privacy. Whereas the freedom of expression protects the right of speakers to disseminate information and for the audience to receive it, the right to privacy includes the protection of secrecy and confidentiality.[380] The protection of such values may also constitute a legitimate aim for restricting the freedom of expression.[381] As both rights are qualified, the interests must be balanced against each other in the particular case. Since protection against the disclosure of private information and defamation aims to safeguard related interests, the legal assessment involves similar aspects. This includes taking into consideration the identity of the victim, the subject matter, the severity of the interference, the contribution to the general interest and the manner in which the information was acquired.[382] For example, publication by the media of photographs taken in public without consent has been considered a violation of the right to privacy, when balancing these rights.[383]

Disclosure of private information may involve the state, such as information shared between different governmental agencies,[384] or non-state actors, be it newspapers[385] or private individuals.[386] Case law on the topic before the ECtHR mainly concerns publications by online newspapers and only to a limited extent by private individuals. As in relation to defamation, given the special role of the press in informing the public, the protection of individuals against disclosure by other private individuals is particularly far-reaching. In terms of the press, the Court has affirmed that the zone of private life is also broader for 'ordinary persons'.[387] Meanwhile, protection of information involving public persons is reduced, as they lay themselves open to public scrutiny, even concerning certain aspects of their private lives.

(18 June 2018), para. 57; CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 38.

[379] *Flinkkilä and Others v. Finland* App no 25576/04 (ECtHR, 6 April 2010), para. 75.

[380] *M.S. v Sweden* App no 74/1996/693/885 (ECtHR, 27 August 1997).

[381] *Editions Plon v France (*2004) 42 EHRR 36 (confidentiality).

[382] *Von Hannover v Germany (No. 2)* (2012) 55 EHRR 15, paras. 108–113.

[383] *Von Hannover v Germany* (ECtHR).

[384] *Z v Finland* (1997) 25 EHRR 371.

[385] *Verlagsgruppe News GmbH and Bobi v Austria* App no 59631/09 (ECtHR, 4 December 2012).

[386] *Volodina v Russia* (ECtHR).

[387] *Sciacca v Italy* (2006) 43 EHRR 400, para. 29.

However, it does not encompass publication merely to satisfy the 'prurient curiosity of a particular readership'.[388]

A few cases involve the disclosure of personal information by another private individual. *Volodina v Russia* concerned the publication of personal photographs on social media without consent, in the context of domestic violence.[389] Although a criminal investigation was initiated, it did not result in prosecution. According to the ECtHR, 'publication of her private photographs further undermined [the victim's]. . . dignity, conveying a message of humiliation and disrespect'.[390] In conjunction with physical violence and psychological abuse, this constituted inhuman treatment.[391] Furthermore, the previously discussed *K.U. v Finland* case involved a situation of doxing, with the publication of a child's personal information in an online dating ad by a private individual. The Court in its assessment affirmed the applicability of Article 8. Even though the offence in question was not as severe as in other cases concerning sexual violence before the Court, for example, involving rape, the act could not '. . .be treated as trivial', as it concerned a minor and made him a target of paedophiles.[392] The negligence of the state to protect the applicant thus violated several aspects of the right to privacy, including the disclosure of private information and the invasion of the victim's sexual autonomy. Positive obligations in this regard involved adopting effective criminal legislation as well as investigating and prosecuting the offender which, as a first step, required the possibility of identifying the perpetrator.

In terms of subject matter, according to the ECtHR, protected personal information relates to those areas that touch upon the core of a person's private life, that is, a content-based view of privacy, rather than a subjective approach to what is intimate.[393] There is a presumption of privacy in relation to certain subject matters, acknowledged by the Court *in casu*. For example, the protection of personal information on a person's sexual life has been affirmed, including nude photographs, allegations of promiscuity and details of extramarital affairs.[394]

Furthermore, case law of both the IACtHR and the ECtHR indicates that it includes protection of one's own image, not to be taken, used or published without consent. [395] The ECtHR has affirmed that both the taking of a photograph of another person and its publication without the consent of the subject violate privacy.[396] In

[388] *Von Hannover v Germany* (ECtHR), para. 43.

[389] *Volodina v Russia* (ECtHR).

[390] ibid., para. 75.

[391] ibid., para. 75.

[392] *K.U. v Finland* (ECtHR), para. 45.

[393] *See*, for example, *Söderman v Sweden* (ECtHR), para. 79.

[394] *Ion Carstea v Romania* (ECtHR); *Biriuk v Lithuania* App no. 23373/03 (ECtHR, 25 November 2008), para. 41; *Somesan and Butiuc v Romania* (ECtHR), para. 26.

[395] *Von Hannover v Germany* (ECtHR), para. 72; *Case of Fontevecchia and D'Amico v Argentina* (merits, reparations, and costs) IACtHR Series C No. 238 (29 November 2011), para. 67.

[396] *Reklos and Davourlis v Greece* App no 1234/05 (ECtHR, 15 January 2009). *See* also *Bogomolova v Russia* App no 13812/09 (ECtHR, 20 June 2017), para. 56.

von Hannover v Germany, involving the publication of photographs in the tabloid press of Caroline von Hannover—a member of the royal family of Monaco—the Court affirmed that states acquire positive obligations to ensure protection of individuals from interference by others, including the publication of photographs of individuals without consent.[397] This requires the adoption of effective civil law legislation and ensuring access to remedies. In this regard, the ECtHR noted the necessity of increased vigilance in protecting people's private life, considering the development of new communications technologies, including the systematic taking of photographs and their widespread dissemination to the public.[398] The special protection of photographs emanates from the fact that such '. . .may contain very personal or even intimate information about an individual and his or her family'.[399] Accordingly, a '. . .person's image constitutes one of the chief attributes of his or her personality, as it reveals the person's unique characteristics and distinguishes the person from his or her peers' and the protection of one's image is '. . .one of the essential components of personal development'.[400] This protection extends to photographs depicting everyday events and not solely to information causing humiliation or embarrassment.[401] Nevertheless, how a person is represented in a photograph is also considered, for example, whether displaying acts of an intimate nature, such as nudity.[402] In *Khadija Ismayilova v Azerbaijan*, the recording and dissemination online of videos displaying a couple's sexual activities was considered particularly grave.[403] As noted above, where the content is of a sexual nature, it may also violate a person's sexual autonomy.

A person's home address is also considered a private matter and may not be published without consent, unless for a public interest.[404] For example, in *Khadija Ismayilova v Azerbaijan*, the authorities had, during the course of a criminal investigation, disclosed personal information such as the applicant's address, the identity of a person with whom she had an extramarital affair and the names and occupations of her friends and colleagues.[405] This data was subsequently published by the press. The Court noted that all these types of information were protected by the right to private life, in this case involving state interference without a legitimate aim.[406]

[397] *Von Hannover v Germany* (ECtHR), para. 72.

[398] ibid., para. 70.

[399] *Verlagsgruppe News GmbH and Bobi v Austria* (ECtHR), para. 66.

[400] ibid., para. 68.

[401] *Von Hannover v Germany* (ECtHR).

[402] *Ion Carstea v Romania* (ECtHR) (nude photographs); *Verlagsgruppe News GmbH and Bobi v Austria* (ECtHR), para. 87 (holding hand over crotch).

[403] *Khadija Ismayilova v Azerbaijan* (ECtHR), para. 116.

[404] *Alkaya v Turkey* App no 42811/06 (ECtHR, 9 October 2012), para. 39.

[405] *Khadija Ismayilova v Azerbaijan*(ECtHR).

[406] ibid., paras. 140–142.

It generally also includes personal health data, in particular a person's status as HIV-positive,[407] information on abortion procedures,[408] and mental health issues.[409] Much case law concerns the disclosure of health-related information, be it by authorities or newspapers. The particularly sensitive and confidential nature of this information has in such cases been highlighted, even more so in instances where the data may cause stigmatisation. For example, in *Armoniene v Lithuania,* the ECtHR held that the publication in a major daily newspaper of information on the HIV status of an individual, in addition to allegations of his having fathered two children out of wedlock, his name and address, was an 'outrageous abuse of press freedom'.[410] The Court reaffirmed the importance of the right to privacy in ensuring the development of a person's personality and that protection extends beyond the family sphere to include a social dimension.[411] This entails that a person's social relationships and standing are issues for consideration. The protection of personal data, such as medical information, was deemed of fundamental importance, including a person's HIV status. Accordingly, the disclosure of such data may '. . .dramatically affect his or her private and family life, as well as the individual's social and employment situation, by exposing that person to opprobrium and the risk of ostracism'.[412] Also, the allegation that he was the father of children by another woman was of a 'purely private nature'.[413] These disclosures caused public humiliation and exclusion from village social life.

The release of reproductive health information has also been considered particularly intrusive. In *P. and S. v Poland*, information from the police and health data were made public through a press release by the hospital, involving the sexual assault of a young girl and her ensuing abortion.[414] This information was subsequently published online in various discussion fora. As the information was disseminated on the Internet, it was accessible to a wider audience. The events were sufficiently descriptive so that others were able to find the applicant's contact details and communicate directly with the girl and members of her family. The Court emphasised the necessity of protecting the right to privacy of individuals, particularly involving details of their sexual life, and the dissemination of such information thus constituted a violation of the right to privacy.[415]

The sphere where the information was acquired or disclosed is a relevant factor, for instance, whether it involves a private or public space, as the right to privacy is

[407] *Armoniene v Lithuania* (2009) 48 EHRR 53, para. 42; *Z v Finland* (ECtHR). It is considered that this type of disclosure may cause stigma.

[408] *M.S. v Sweden* (ECtHR).

[409] *Panteleyenko v Ukraine* App no 11901/02 (ECtHR, 29 June 2006).

[410] *Armoniene v Lithuania* (ECtHR), para. 47.

[411] ibid., para. 39.

[412] ibid., para. 40.

[413] ibid., para. 42.

[414] *P. and S. v Poland* App no 57375 (ECtHR, 30 October 2012).

[415] ibid., paras. 133–134.

traditionally limited to the former. The categorisation of the Internet is thus of particular relevance. As mentioned in Chap. 3, although the Internet mainly has characteristics of a public sphere, it contains private pockets, such as private messaging functions. This delineation has an impact on the scope of privacy. In early case law, the European Commission affirmed that taking photographs without the consent of the person may, under certain circumstances, constitute an intrusion of a person's privacy, such as taking photographs in a person's home without consent.[416] This also extends to situations where the photograph was taken with consent in a person's home, with an expectation of it remaining private, but later disseminated to the public.[417] In such instances, there has been an intrusion of the "inner circle" of the person's private life, as opposed to photographs taken in public.[418] In more recent case law, the Court has rejected the separation of spatial spheres—the public and the secluded—as being too vague and indeterminate.[419] Rather, in *Halford v the United Kingdom*, the ECtHR affirmed that individuals also maintain a degree of privacy in public spaces, although not as extensive as in private, relative to an individual's "reasonable expectation of privacy".[420] Since there are occasions when people knowingly or intentionally involve themselves in activities which may be recorded or conveyed in a public manner, a person's reasonable expectation of privacy is a significant, albeit not necessarily conclusive, factor.[421] This supports a content-related rather than spatial division of public and private spheres. Privacy in the public sphere also extends to social interactions. As the right to privacy includes the development, without outside interference, of the personality of each individual in his/her relationships with other human beings, there is a '. . .zone of interaction of a person with others, even in a public context, which may fall within the scope of "private life"'.[422]

As noted in relation to defamation, the size of the audience is also relevant in the assessment of harm.[423] Whether the information was published on a news website, blog or social media and the number of its monthly viewers is thus relevant. Furthermore, the nature of the website in question is relevant from the perspective of non-consent. As the offence involves solely *non-consensual* dissemination, whether a person has consented to the release of information must thus be evaluated. Particularly in this regard, the Internet challenges theoretical presumptions concerning the protection of secrecy, discussed in Sect. 3.3.3. It requires a clear approach to whether consent may be presumed when information is shared with a

[416] *Friedl v Austria* App no 15225/89 (Commission Decision, 19 May 1994), paras. 49–50.

[417] *Verlagsgruppe News GmbH and Bobi v Austria* (ECtHR), para. 84.

[418] *Friedl v Austria* (ECtHR), para. 49.

[419] *Von Hannover v Germany* (ECtHR), para. 75.

[420] *Halford v the United Kingdom* (1997) 24 EHRR 523, para. 45.

[421] *P.G. and J.H. v the United Kingdom* (2008) 46 EHRR 51, para. 57.

[422] *Von Hannover v Germany* (ECtHR), paras. 50, 69.

[423] *Karhuvaara and Iltalehti v Finland* App no 53678/00 (ECtHR 16 November 2004), para. 47; *Von Hannover v Germany (No. 2)* (ECtHR), para. 112.

smaller audience, or whether a person has a right to control information that has already been released, that is, removing the focus on complete secrecy as determinative of whether an invasion has occurred. This has, to an extent, been touched upon in cases on the republication of information. The general approach to secrecy of the ECtHR is that with information already available on the Internet, the protection of confidentiality is reduced. For example, the case of *Editions Plon v France* concerned the publication of a book, detailing medical information on former President Mitterrand, provided by his physician.[424] A ban on distribution was issued as the book allegedly breached medical confidentiality. However, many copies had already been sold by that time and information from the book had been published in articles and on the Internet. Thus, the information in the book was to a large extent no longer secret in practice and the preservation of confidentiality was not an overriding interest. There was accordingly no longer a "pressing social need" to ban the book.[425]

Nevertheless, a degree of protection remains in certain instances of republication. In *Aleksey Ovchinnikov v Russia*, the reposting of information on the Internet, which had initially been printed in two news articles, was assessed.[426] The printed news articles described a violent incident, including sexual assault, perpetrated by several 12-year-olds. In the second article, the alleged juvenile offenders were named and information about their relatives was included, as these held prominent positions in Russia. By the time of the second publication, the information was already available online and in an article in another newspaper. The information was based on the result of official investigations and was thus, by all accounts, true. It was therefore not a matter of defamation. The journalist was, however, held liable for disclosing private information about the offenders and their families. The Court noted that, by the time of the second publication, the offenders' personal information was no longer confidential and already in the public domain. Accordingly, the necessity of '. . .protecting the identity of the juvenile offenders and their relatives had been substantially diminished, so that the preservation of confidentiality in this matter could no longer constitute an overriding requirement'.[427] However, the fact that the information had already entered the public domain did not necessarily entail that a restriction on reproducing information was not justified, to prevent '. . .further airing the details of an individual's private life which do not come within the scope of any political or public debate on a matter of general importance'.[428] This correlates with theories on limited privacy protection, that is, extending the right to privacy to situations where information, albeit previously disclosed, is distributed beyond the previous network without consent. It thus affirms a right to a degree of privacy beyond secrecy.

[424] *Editions Plon v France* (ECtHR).

[425] ibid., para. 53.

[426] *Aleksey Ovchinnikov v Russia* (ECtHR).

[427] ibid., para. 49.

[428] ibid., para. 50.

It should be noted that this case involved a balancing between the right to privacy and the freedom of expression of journalists, which is broader than for private individuals. The Court underscored that in instances where details of an individual's private life are published for the sole purpose of satisfying the curiosity of a particular readership, the individual's private life prevails over the freedom of expression of journalists.[429] The reposting of personal information by private individuals would thus receive even less protection. In this case, the information on a minor involved in a criminal act made no contribution to the public interest, with the Court noting the more extensive right to effective protection of the private lives of minors.[430] The publication of this information was deemed harmful to the boy's moral and psychological development and his private life.[431] Civil liability was thus deemed proportionate. The vulnerable position of children, to be taken into account in the balance between the freedom of expression and the right to privacy, has also been emphasised in other cases.[432]

The main venue for accountability is civil claims against the individual, media or authority disclosing the information. In certain instances—involving a risk of sexual violence or dissemination of sexual content—effective criminal law remedies have been required. None of the cases concern intermediary liability. However, as the legal assessment of, for example, the proportionality of restrictions is similar as *vis-à-vis* defamation, it is likely that the approach to intermediary and media publisher liability would be comparable. The most reasonable approach, employed in multiple states, is a notice-and-takedown system, which may be in the form of a notice-wait-and-takedown, allowing for the possibility of counter notices.[433] Nevertheless, this is not construed as an obligation under international human rights law or EU law, apart from the removal due to requests under the right to be forgotten. It is particularly challenging for AI and human moderators to assess the consent of individuals and the veracity of statements. Nevertheless, as discussed previously, intermediaries are increasingly using machine learning to detect and automatically flag intimate images shared without consent. This is either subject to domestic law or corporate codes of conduct, or on their own initiatives. As noted, where disclosure falls within the definition of "cyber harassment", the proposed EU directive also obliges states to ensure removals or blocking of content by intermediaries upon orders by judicial authorities.[434] At a general level, online platforms are also

[429] ibid., para. 50.

[430] ibid., para. 51.

[431] ibid. para. 51.

[432] *Kurier Zeitungsverlag und Druckerei GmbH v. Austria* (No. 2) App no. 1593/06 (ECtHR, 19 June 2012), para. 59; *K.U. v Finland* (ECtHR), para. 41.

[433] *See*, for example, New Zealand, the Harmful Digital Communications Act 2015, Public Act 2015 No 63, Section 24.

[434] Art. 25 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

developing privacy-enhancing tools and website designs for easier navigation, reducing the risks of disclosures.

4.3.5.3 Conclusion

The values to be protected through the prohibition of the disclosure of information without consent is mainly that of personal development and integrity, as each person has a right to a zone of privacy free from display to others. The protection of secrecy thus extends to personal information in general. The zone of privacy is in these cases not viewed as spatial, but rather content-based. This places positive obligations on states to protect the reputation of individuals also in such public spheres as the Internet. However, the space is relevant in assessing the scope of privacy. For example, although the ECtHR has affirmed that a zone of privacy exists also in public, the scope is determined in relation to what may be reasonably expected. As the Internet may be categorised as a public sphere with private pockets, it requires a contextual approach to the scope of privacy. Issues of consent, confidentiality and reasonable expectations of privacy may be construed differently online.

Even if the delineation of protected subject matter is linked to social development, the ECtHR clearly holds that information related to sexuality, health, family relations, criminal behaviour and contact information is included in the private sphere, as are private photographs and videos. As noted by the Court, the right to privacy involves a social dimension, which encompasses the interaction between individuals as well as a person's public image. The subject matter is thus in this regard objectively rather than subjectively determined. It is not a requirement that the information causes humiliation or is particularly sensitive. As noted above, the publication of photographs of a person running errands was considered a violation of the right to privacy (*von Hannover v Germany*). However, depending on the content of the information disclosed, additional harm may arise, for example, in relation to a person's social standing. This particularly relates to facts deemed to cause humiliation or stigma. Such an approach has been taken by the ECtHR in cases involving a person's HIV status, reproductive health information and sex life. As the consequences of the publication for the individual is of importance, the content of the material is relevant in assessing the degree of harm, both concerning the subject matter—with photographs of a sexual nature deemed particularly intimate—and how the person is represented.

Gender-based disclosures frequently involve intimate photographs and information related to sexuality, health and contact details, which have all been affirmed as protected areas within the right to privacy. Since these areas are objectively determined as aspects of a person's private life, they are gender-neutral. Nevertheless, the question may be raised whether there is a risk in interpreting the scope of the protection of reputation as relative to social condemnation, for example, a person's honour. It may produce gendered approaches to dignitary harms, such as stereotypical perceptions of female sexuality. Given that a particular stigma is considered attached to certain subject matters, such as sexuality and specific medical conditions,

select aspects of a person's life may be categorised as particularly shameful. It cannot be presumed that certain types of subject matter are categorically private, particularly in light of current practices of sharing personal information online. Rather, non-consent is central. At the same time, given that gender stereotyping commonly involves the shaming of female sexuality, sexual objectification and prescriptive norms on modesty, certain forms of disclosure may be particularly harmful to women.

Additionally, the ECtHR in its case law emphasised that photographs require especially extensive protection as they are an important attribute of a person's personality. This entails that the publication of intimate photographs of private individuals without consent clearly violates the right to privacy. The publication of content of a sexual nature may also constitute a form of sexual violence. Whether a photograph has been taken without the consent of the person in question has been a consideration of the ECtHR, as an additional element in assessing harm and the person's reasonable expectations of privacy. Implied is a presumption of non-consent to the subsequent disclosure. Naturally, it may be considered particularly intrusive when both the taking of the photograph and its publication are non-consensual. However, the fact that a photograph was taken consensually should not considerably affect the evaluation of harm, since consent to the taking of a photograph may be dependent on it remaining in the private sphere. This was indicated in *Verlagsgruppe News GmbH and Bobi v Austria*, where a photograph was taken in the home, clearly with a view of keeping it private. As mentioned above, in most cases of so-called revenge pornography, the photographs have been taken by the victim or by someone else with consent, with the non-consensual distribution still causing substantial harm.

The notion of information intended for the private sphere deserves particular consideration in the context of the Internet. It should be noted that the ECtHR in *Aleksey Ovchinnikov v Russia* held that the scope of privacy is reduced once confidentiality is lost, for example, through publication on the Internet. Confidentiality is rare in the Internet age, thus engendering more limited protection if the same legal approach is upheld as in other media. However, the Court held that a certain degree of protection of private information remains, although it is no longer confidential, in view of a person's reputation and dignity. This is especially the case concerning children. Whether and how this approach extends to specific forms of dissemination on the Internet is unclear, for example, if it involves social media. For instance, is information and photographs disclosed to a small group of individuals no longer confidential? Would a person in such situations maintain a reasonable expectation of privacy, for example, by taking steps to restrict access through his/her privacy settings? As discussed previously, various theories have been developed in order to assess the scope of limited privacy, where information is not secret but may still warrant protection. This includes theories on common social patterns of distributing information as well as user restrictions on access to information. Such theories do not entail a strictly objective delineation of certain areas as inherently private as this may vary, particularly in relation to sexual practices through new technologies. From such a theoretical standpoint, expectations of privacy may be

reasonable even when a person shares information with a smaller audience on social media, if actively restricting access by using privacy settings. It may also encompass situations where a person takes intimate photographs with consent and sends them to another person, with the clear expectation of them remaining in that limited setting.

Additionally, the delineation of the scope of privacy is assessed in relation to several factors connected to the public/private distinction. Private, as opposed to, public individuals are entitled to broader protection of their privacy as they have not chosen to be in the public spotlight. The demarcation between public and private persons is not fixed. However, it is clear that the latter category includes politicians and entertainers, whereas certain professionals, such as lawyers, journalists and university professors, mainly have been considered to be private persons, while public figures if well known.[435] As noted in relation to defamation, such distinctions are particularly challenging in the context of the Internet where private individuals may become known to larger audiences, for instance, through social media. The scope of the person's Internet presence is thus relevant. Nevertheless, even public figures have a right to a private life, also in the public sphere, as the Court clearly distinguishes between public and personal information related to these individuals. In relation to the media, this is correlated with whether the information is of a general interest or merely information for the purpose of satisfying the curiosity of the general public. The importance of the press as a public watchdog is in this regard emphasised, involving publications in both print and on the Internet. From this perspective, the right of the audience to receive information, encompassed in democracy theories on the freedom of expression, entails that the personal interest of restricting access to information may be overridden, for example, where disclosures are made in the media involving women in positions of authority.

In the case law discussed in this section, personal information was generally disclosed by authorities and newspapers. Few cases concerned the publication of personal information by private individuals, which is the most common form of gender-based disclosure online. However, it is clear that where the victim and perpetrator are private individuals, the protection is particularly extensive. *K.U v Finland* involved doxing, with the ECtHR clearly affirming state obligations to protect individuals against sexual violence, considering the elevated risk of abuse when contact details are published on the Internet. In *Volodina v Russia*, a former partner published photographs online as a form of domestic violence and the disclosure was thus analysed in that context, constituting inhuman treatment in conjunction with other acts.

In terms of positive obligations *vis-à-vis* the right to privacy, the ECtHR in *von Hannover v Germany* required the adoption of effective civil legislation and procedures. Meanwhile, in *K.U v Finland* and *Volodina v Russia*, the Court affirmed obligations to ensure effective criminal laws and prosecution, given the severity of the offences, being sexual- and domestic violence. The proposed EU directive on

[435] *Ion Carstea v Romania* (ECtHR), para. 37; *Khadija Ismayilova v Azerbaijan* (ECtHR), para. 119.

cyber violence also requires the criminalisation of cyber harassment. It is thus clear that the protection against disclosure in international human rights law generally translates to the Internet, delineating state obligations against various forms of gender-based disclosures. However, obligations have not as of yet been affirmed to encompass regulation of secondary liability for intermediaries in relation to this form of content, beyond general statements involving image-based sexual abuse, and the anticipated EU directive encompassing online harassment. As noted, the preferred mechanisms at the domestic level are versions of notice-and-takedown systems. Given the challenges that arise in assessing consent in relation to published materials, this is the most reasonable venue concerning most forms of disclosure, plausibly with more extensive obligations involving image-based sexual abuse. The development of AI that require or assess the consent of published material, or privacy-enhancing tools and website design, are also means of preventing gender-based harm.

4.4 Hate Speech

4.4.1 Theorising Harm

The following section provides an overview of the approach to hate speech in international human rights law. Currently, this field of law does not explicitly extend the concept of hate speech to "sex" or "gender" as protected grounds. The possibilities—and the suitability—of expanding the legal framework on hate speech to include derogatory speech against women as a group will thus be explored, given its widespread occurrence online.

The rationales for allowing content-based restrictions on expressions, such as hate speech, are multiple from a theoretical perspective. As noted in Sect. 3.2 on harm, domestic laws and legal theories to varying degrees accept harm emanating from speech, be it individual or social, viewed also in relation to hate speech. In instances where domestic laws prohibit hate speech from the perspective of the individual, they frequently aim to protect a person's health, autonomy, security, human dignity and participation in democratic dialogue.[436] Individual harm resulting from hate speech is thus generally understood as psychological, including the internalisation of stereotypes, self-defeating attitudes or antisocial behaviour,[437] often addressed through domestic laws on tort of intentional infliction of emotional distress. The latter often requires proof of psychological harm in the individual case, which may not arise in all instances of hate speech. [438]

[436] Brown (2015), p. 1.

[437] Bennett (2016), p. 474; Brown (2017), p. 27.

[438] Bennett (2016), p. 487.

In contrast, social and group-based harm is central to the offence of hate speech. Even though such speech may be directed either at a group or an individual, the individual is targeted on the basis of his/her membership in a particularly protected group. Accordingly, individual harm is aggregated into broader structures, causing or contributing to social harm of such severity that regulation is warranted.[439] Social harm includes the reinforcement of societal and institutional prejudice, depriving groups of rights and power, the enactment and perpetuation of subordination, the encouragement of social norms that impede effective counter-speech and violence.[440] This is asserted primarily from a theoretical standpoint. For example, according to linguistic theories, hate speech is a form of speech act, by some categorised as illocutionary speech, that is, not merely producing discrimination and social hierarchies, but *constituting* discrimination.[441] As such, 'the category of fighting words reverses the dialectics between speech and conduct: speech which can provoke retaliatory conduct is in itself a form of conduct'.[442] This has been affirmed by, *inter alia*, the CoE Parliamentary Assembly, which considers racism to be a criminal act rather than an opinion.[443] If speech is an act, harm occurs through the very act of speaking. As speech and its effects are synonymous in relation to the concept of illocutionary speech, empirical assessments of its impact are not necessarily required.[444] Hate speech may also be categorised as perlocutionary speech, understood as speech creating a climate conducive to harmful acts. Such speech increases economic, political and social inequalities as well as the risk of hate crimes, such as violence against a particular group.[445] Arguably, hate crimes rarely occur without the prior stigmatisation and dehumanisation of a group and a contextual approach is thus necessary.[446]

The viewpoint that hate speech is harmful at a social level consequently assumes that it constitutes or contributes to inequality. However, the cause and effect of speech and social harm, such as hostility, discrimination and violence, is contested.[447] This is in part related to the fact that few states collect data on hate crimes and their causes.[448] Certain empirical studies in psychology and criminology affirm a correlation, such as the development of negative stereotypes as a result of

[439] ibid., p. 446.

[440] Brown (2017), p. 31.

[441] Matsuda (1989), p. 2358.

[442] Rigaux (2004), p. 292. *See* also Tulkens (2012), p. 281.

[443] CoE Parliamentary Assembly, 'Recommendation No. 1543: Racism and xenophobia in cyberspace' (2001), para. 1.

[444] Levin 2010), p. 113.y

[445] Strossen (1996), p. 449.

[446] UN HRC, 'Report of the Special Rapporteur on Minority Issues, Rita Izsák' (2 January 2015) UN Doc A/HRC/28/64, para. 26.

[447] Gagliardone et al. for UNESCO (2015), pp. 35, 54.

[448] UN HRC, 'Report of the Special Rapporteur on Minority Issues, Rita Izsák' (2 January 2015), para. 26.

content in the media or in video games.[449] However, in general, there is limited research on the contribution of speech to climates of hatred and hate-based discrimination and violence.[450] Arguably, the causes of social inequalities and harm may arise from culture, value preferences and lifestyle patterns independent of hate speech, and it may be inaccurate to attribute social harm to speech.[451] At the same time, while many states require causality between conduct and harm in order to criminalise behaviour, this is not necessarily the case in relation to hate speech laws, as there is no principal offence—that is, hatred. Instead, it involves the prevention of exacerbated risks of social harm.[452] It should also be borne in mind that the same level of causality is not required in international human rights law.

The viewpoint that hate speech generates inequality has also been addressed as an aspect of the freedom of expression. Free speech is integral to such values as democracy and individual autonomy. Meanwhile, these ideals simultaneously involve the protection of vulnerable communities.[453] As noted above, a substantive equality approach *vis-à-vis* the freedom of expression and democracy requires that everyone should be guaranteed equal access to communication and participation in debate. This places a demand on the regulation of harmful speech demeaning vulnerable groups in society, from an access-oriented as well as a substantive approach, in order to make room for marginalised individuals and multiple perspectives in public discussions. Accordingly, as argued by Owen Fiss, the regulation of hate speech should not be construed as a conflict between non-discrimination and the freedom of expression, involving two different norms. Rather, it involves competing claims of the freedom of expression, in relation to the rights of audiences and speakers.[454] From a legal theory perspective, it can be argued that hate speech has the effect of silencing the targeted group, which requires state interference in order to ensure that all sides are able to participate and be represented. Accordingly, '[s]ometimes we must lower the voices of some in order to hear the voices of others'.[455] Nevertheless, regardless of whether construed as a conflict between rights, intra-rights or between groups, a balancing of interests and values ensues.

Furthermore, even if individual or social harm is recognised, there is disagreement as to the measures that are able to constrain such speech, related to the concept of harm. Certain scholars consider that the social benefits of protecting speech outweigh the harm, in line with, for example, the marketplace of ideas theory. That is, hate speech is considered harmful but the harm is not sufficiently severe to warrant regulation. It is arguably useful to allow speech deemed offensive. As such,

[449] Burgess et al. (2011); Williams et al. (2020). *See* also Brown (2015), p. 69.

[450] Brown (2015), p. 70.

[451] Bennett (2016), p. 445.

[452] Brown (2015), p. 69.

[453] 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, para. 4.

[454] Fiss (1996), p. 26. *See* also Mahoney (2010), p. 97; Elbahtimy (2014), p. 7.

[455] Fiss (1996), p. 18.

'[t]he most obnoxious expressions of racism confront us with the offensiveness of the speakers and their ideas. We obtain an important truth from these speakers, although it is not the truth they mean to convey'.[456] Societies evolve through disagreement and contestation.[457] Additionally, legal suppression of such speech may pre-empt superior non-legal responses to offensive speech.[458] Scholars have, for example, contested the efficacy of Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination (the ICERD)—prohibiting incitement to racial discrimination—as not being empirically proven.[459] Social policies for inclusion rather than criminal laws on speech are preferred.[460] There are also broader risks associated with prohibiting such speech. The content of hate speech legislation runs the risk of being subjective, as it '. . .tends to be the expression of the dominant group that controls the content of the law'.[461] Accordingly, '[u]nderlying all attempts to practice viewpoint discrimination is a belief in an identifiable and preferable social norm'.[462]

In contrast, it has been argued that hostile expressions do not contribute to the free flow of ideas,[463] the empirical link between allowing hate speech and the defusing of racial hatred or homophobia is elusive[464] and the burden of the harmful effects of such speech are borne by vulnerable groups.[465] There is thus a clear cultural and theoretical divide among states in relation to hate speech, apparent also in the codification process of the Convention on Cybercrime of the Council of Europe (the Budapest Convention), which occasioned disagreement over the inclusion of hate speech.[466] Consequently, although racist and sexist speech is generally considered harmful, both the asserted effect of hate speech on a social level and the necessity of its criminalisation are disputed.

[456] Farber (1980), p. 301.

[457] Gagliardone et al. for UNESCO (2015), p. 15.

[458] Goldberger (1991), p. 1209.

[459] Sandmann (1994), p. 251.

[460] Gagliardone et al. for UNESCO (2015), p. 57.

[461] ibid., p. 54. *See*, for instance, the dissenting opinion of Judge András Sajó, joined by Judges Vladimiro Zagrebelsky and Nona Tsotsoria in *Féret v Belgium* App no 15615/07 (ECtHR, 16 July 2009).

[462] Sandmann (1994), p. 252.

[463] Tsesis (2002), para. 30.

[464] Delgado and Stefancic (2018), p. 205.

[465] Waldron (2012), p. 5.

[466] This resulted in an additional protocol, rather than inclusion in the main treaty.

4.4.2 *International Human Rights Law*

In international human rights law, certain provisions *allow* states to restrict expressions, whereas others *oblige* states to prohibit harmful speech. In relation to the former, states may restrict such speech, for example, for the purpose of protecting the rights and freedoms of others, which extends to both individuals and communities.[467] This also pertains to groups not included in the scope of hate speech, provided the requirements of freedom of expression provisions are met, such as the pursuit of a legitimate aim. Meanwhile, obligations to prohibit hate speech can be found in the ICERD,[468] the ICCPR,[469] the ACHR[470] and the Additional Protocol to the Budapest Convention.[471] For example, while Article 19 of the ICCPR *permits* states to restrict expressions, Article 20 of the same Convention *obliges* states to prohibit '[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. . .'.[472] Certain treaties also include provisions omitting protection for activities aimed at the destruction of the rights and freedoms of the convention in question, which encompasses hate speech.[473]

International human rights law primarily approaches hate speech in terms of social harm, while also providing for the adjudication of individual cases. Even though conventions such as the ICERD distinguish between racial discrimination and racist hate speech,[474] the underlying reasons for international provisions on hate speech are generally to ensure non-discrimination, the dignity and safety of members of the group as well as social peace and political stability.[475] Hate speech is

[467] *Malcolm Ross v Canada* (UNHRC), para. 11.5.

[468] Art. 4 (a) of International Convention on the Elimination of All Forms of Racial Discrimination (1965), 21 December 1965, GA res 2106 (xx), Annex, 20 UN GAOR Supp. (No. 14) at 47, UN Doc. A/6014 (1966), 660 UNTS 195, entered into force 4 January 1969.

[469] Art. 20 (2) of the ICCPR.

[470] Art. 13 (5) of the ACHR.

[471] Chapter II, Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems, (ETS No. 189) 28 January 2003.

[472] Similarly, Art. 13 (2) of the ACHR allows states to restrict speech to ensure certain public interests, whereas Art. 13 (5) obliges states to prohibit speech that incites violence against certain groups.

[473] Art. 5 of the ICCPR; Art. 17 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), 4 November 1950, 213 UNTS 222, entered into force 3 September 1953. UN HRC has only applied Art. 5 (1) in *M.A. v Italy*, Communication No. 117/1981, UN HRC, UN Doc. A/39/40 (10 April 1984), concerning the reorganising of a dissolved fascist party.

[474] Art. 1 and Art. 4 of the ICERD.

[475] The IACmHR holds that the purpose of the prohibition on incitement to lawless action is the protection of dignity and non-discrimination. *See* IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas', Annual Report of the Office of the Special Rapporteur for Freedom of Expression: Annual Report of the Inter-American Commission on Human Rights Vol. II (31 December 2015) OEA/Ser.L/V/II, para. 18; *Féret v Belgium* App no 15615/07 (ECtHR, 16 July 2009), para. 73. According to the ECtHR,

commonly understood as involving incitement to discrimination, in addition to, for example, violence.[476] It also undermines access for the targeted group to specific rights and freedoms, such as the freedom of expression.[477] An equality perspective *vis-à-vis* restrictions on the freedom of expression is thus considered valid and the two rights are understood as 'mutually supportive'.[478] For example, the United Nations Committee on the Elimination of Racial Discrimination (CERD) construes its obligation to prohibit hate speech as a means of ensuring equality *vis-à-vis* the freedom of expression, as racist speech 'potentially silences the free speech of its victims'.[479] This in turn is linked to the overarching protection of democracy. As noted by the ECtHR, democracy does not solely entail that the views of the majority prevail but that a balance is achieved that ensures the proper treatment of minorities and avoids the abuse of dominant positions.[480]

The ECtHR has also acknowledged the protection of individual dignity, affirming that states incur positive obligations to prohibit speech in order to protect individuals from group-based negative stereotyping, as a means of ensuring the right to privacy. Accordingly, ethnicity is considered an aspect of the social identity of individuals and '. . .any negative stereotyping of a group, when it reaches a certain level, is capable of impacting on the group's sense of identity and the feelings of self-worth and self-confidence of members of the group. It is in this sense that it can be seen as affecting the private life of members of the group'.[481] Similar reasoning has been offered in a case concerning hate speech on social media *vis-à-vis* a same-sex couple, with the ECtHR holding that it affected the psychological well-being and dignity of the applicants.[482]

---

such speech undermines the dignity and the safety of certain groups and represents a danger for social peace and political stability in democratic States. *See* also UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 4.

[476] *See*, for example, Art. 20 (2) of the ICCPR; Art. 2 (1) of the Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems (ETS No. 189) 28 January 2003.

[477] The CERD Committee emphasises that the freedom of expression '. . .should not aim at the destruction of the rights and freedoms of others, including the right to equality and non-discrimination.' *See* CERD, 'General Recommendation No. 35: Combating Racist Hate Speech' (26 September 2013) UN Doc. CERD/C/GC/35, para. 26; *Norwood v the United Kingdom* (2005) 40 EHRR SE11.

[478] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (7 September 2012) UN Doc. A/67/357, para. 3. *See* also UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 4; CERD, 'General Recommendation No. 35: Combating Racist Hate Speech' (26 September 2013), para. 45.

[479] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', paras. 28, 45.

[480] *Beizaras and Levickas v Lithuania* (ECtHR), para. 106.

[481] *Aksu v Turkey* (ECtHR), para. 58.

[482] *Beizaras and Levickas v Lithuania* (ECtHR), para. 117.

Nevertheless, it is clear that the prohibition on hate speech must be viewed as an exceptional measure. Several international bodies have noted the value of countering such ideas, and that other measures, for instance, education, may be more reasonable when obligations to combat hate speech can be said to exist.[483] The ECtHR has also argued in favour of restraint in prohibiting expressions categorised as hate speech, emphasising the exclusive use of regulation, albeit acknowledging the necessity of sanctions in certain instances.[484] Accordingly, '. . .undisguised calls on attack on. . .applicants' physical and mental integrity' require protection through criminal law.[485]

There is no coherent and explicit definition of hate speech in international human rights law treaties. Rather, certain provisions involve acts generally understood to be encompassed by the concept. For example, the Additional Protocol to the Budapest Convention obliges states to prohibit '. . .any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, colour, descent or national or ethnic origin, as well as religion if used as a pretext for any of these factors'.[486] Meanwhile, the ICERD and the ICCPR delineate the offences of dissemination of racist propaganda, and incitement and advocacy of hatred.[487] Certain soft law documents also contain definitions.[488] In contrast, such has not been developed by regional human rights law courts in their jurisprudence. As noted by UNESCO, it is unlikely that a universally shared

[483] IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas' (31 December 2015), para. 21. The IACmHR considers that hate speech laws are not sufficient in contexts of structural social inequalities and prejudice, and measures must include, for example, education on harmful stereotypes. As observed by the Special Rapporteurship on Freedom of Expression of the IACmHR, 'speech that offends because of the intrinsic falseness of its racist and discriminatory content must be refuted, not silenced: those who promote these points of view need to be persuaded of their error in public debate.' *See* IACmHR, 'Annual Report of the Office of the Special Rapporteur for Freedom of Expression, Chapter II (Evaluation of the State of Freedom of Expression in the Hemisphere)' (4 March 2011), OEA/Ser.L/V/II. Doc. 5, para. 50. *See* also UN HRC, 'Report of the Special Rapporteur on Minority Issues, Rita Izsák' (2 January 2015), para. 60; UNGA, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (10 August 2011) UN Doc. A/66/290, para. 83.

[484] *Gündüz v Turkey* (2005) 41 EHRR 5, para. 40; *Stomakhin v Russia* App no. 52273/07 (ECtHR, 9 May 2018), para. 117.

[485] *Beizaras and Levickas v Lithuania* (ECtHR), para. 128.

[486] Art. 2 (1) of the Additional Protocol to the Budapest Convention.

[487] Art. 4 of ICERD and Art. 20 (1) of the ICCPR.

[488] *See*, for example, ECRI General Policy Recommendation No. 15 on Combating Hate Speech (adopted on 8 December 2015), p. 3; CoE, Recommendation No. R(97)20 of the Committee of Ministers to Member States on "Hate Speech" Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies), p. 107; UN HRC, 'Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that Constitutes Incitement to Discrimination, Hostility or Violence', Appendix (Rabat Plan of Action) (11 January 2013) UN Doc. A/HRC/22/17/Add.4.

definition will develop due to the contested nature of its regulation.[489] The lack of a clear delineation of the offence entails a risk that states abuse the opportunity to restrict speech,[490] while also allowing for an evolutive approach to its scope, responsive to social change.

Nevertheless, certain common elements may be gleaned from the sources mentioned. In relation to protected groups, provisions in the ICERD, the ICCPR and the Additional Protocol to the Budapest Convention focus on racial, ethnic or national groups.[491] Other provisions are general but have in practice also mainly been applied to the same groups.[492] An individual is targeted as a member of such a group, with the knowledge or belief by the perpetrator of that membership. No explanation is provided why these protected grounds are in focus. However, disparaging speech against groups is not considered in isolation, but rather in its social context, such as historical discrimination and animosity against the group in question. Given the background of the Holocaust as the main catalyst for the development of an international legal framework for human rights, the protection of racial and ethnic minorities has been a prominent consideration. Such speech arguably constitutes a particularly grave threat to social cohesion.[493] The CERD has, for example, provided that the purpose of the ICERD is to combat speech contributing to the creation of a climate of racial hatred, discrimination and the revival of authoritarian ideologies.[494] Meanwhile, the ECtHR has referred to Article 4 (a) of the ICERD in certain instances, which places an obligation on states to make the dissemination of ideas based on racial hatred an offence.[495] Likewise, it has referenced the Recommendation of the Committee of Ministers on hate speech of 1997, which similarly focuses on racist and anti-semitic speech.[496] As such, the ECtHR has connected its interpretation of the freedom of expression to such international obligations, thus also focusing on racial or ethnic hatred.[497]

[489] Gagliardone et al. for UNESCO (2015), p. 8.

[490] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 1.

[491] Art. 20 (2) of the ICCPR; Art. 1 of the ICERD; Art. 2 (1) of the Additional Protocol to the Budapest Convention. *See* also Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law (2008) OJ L328/55 (also religious groups).

[492] For example, Art. 17 of the ECHR.

[493] Brennan (2009), p. 128.

[494] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 5; CERD, 'General Recommendation XV on Article 4 of the Convention', 42nd session (1993), para. 1.

[495] *Jersild v Denmark* (1995) 19 EHRR 1, paras. 30–31.

[496] Recommendation No. R(97)20 of the Committee of Ministers to Member States on "Hate Speech" Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies), p. 107.

[497] For example, in *Jersild v Denmark*, para. 30, the Court held that '[t]he object and purpose pursued by the UN Convention [CERD] are of great weight in determining whether the applicant's conviction which – as the government have stressed – was based on a provision enacted in order to ensure Denmark's compliance with the UN Convention, was necessary within the meaning of

Nevertheless, the ECtHR has extended its approach to hate speech to encompass religious groups.[498] Moreover, in *Vejdeland and Others v Sweden*, the Court broadened its protection to include sexual orientation as a protected ground.[499] Although not explicitly categorising the material in question as "hate speech", it has been affirmed in subsequent case law.[500] The ECtHR in the judgment argued that although its approach to hate speech so far had been limited to cases of racism, xenophobia and religious hatred, sexual orientation should be treated in the same manner.[501] Concurring Judge Spielmann and Judge Nussberger additionally considered the vulnerability of the LGBTQ community in relation to stereotypes and discrimination in Europe.[502] This contextual element was, however, not explored by the majority. The IACmHR has also called for the inclusion of sexual orientation, gender identity, or bodily diversity as protected grounds.[503]

In contrast, the ECtHR did not extend the scope of protected groups in *Savva Terentyev v Russia*, when assessing vulgar speech directed at the police.[504] The Court held that the police '. . .can hardly be described as an unprotected minority or group that has a history of oppression or inequality, or that faces deep-rooted prejudices, hostility and discrimination, or that is vulnerable for some other reason, and thus may, in principle, need a heightened protection from attacks committed by insult, holding up to ridicule or slander. . .'.[505] Although this section implies that hate speech provisions should mainly encompass vulnerable groups, there is no automatic correlation between vulnerability and protected groups, as the case law of the ECtHR has affirmed the vulnerability of a range of groups that have not been recognised as protected from hate speech, such as children,[506] asylum seekers[507] and people with disabilities.[508] Nevertheless, vulnerability in terms of a historical, socio-economic disadvantage appears to be an element of protection against hate speech. In instances of negative stereotyping of a group addressed through the right to

Article 10 paragraph 2 (. . .) Denmark's obligations under Article 10 must be interpreted (. . .) so as to be reconcilable with its obligations under the UN Convention'.

[498] *Norwood v the United Kingdom* (ECtHR).

[499] *Vejdeland and Others v Sweden* (2014) 58 EHRR 15. It is unclear whether the majority was in fact acknowledging that the words constituted hate speech or merely serious and prejudicial allegations. Concurring Judge Yudivska, joined by Judge Villiger, criticised the majority for not explicitly categorising the speech as hate speech.

[500] *Beizaras and Levickas v Lithuania* (ECtHR), para. 125; *Carl Jóhann Lilliendahl v Iceland* App no 29297/18 (ECtHR, 12 May 2020), para. 39.

[501] *Vejdeland and Others v Sweden* (ECtHR), para. 55.

[502] ibid., Concurring opinion of Judge Spielmann joined by Judge Nussberger, para. 6.

[503] IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas' (31 December 2015), para. 13.

[504] *Savva Terentyev v Russia* (ECtHR).

[505] ibid., para. 76.

[506] *A v the United Kingdom* (1999) 27 EHRR 611, para. 22.

[507] *M.S.S. v Belgium and Greece* (2011) 53 EHRR 2, para. 251.

[508] *Alajos Kiss v Hungary* (2013) 56 EHRR 38, para. 42.

privacy, it has also been applied to homogenous groups beyond those protected against hate speech, such as survivors of the Holocaust.[509]

As for the types of speech included, there is a common core in international treaties in prohibiting speech that may cause violence. The intended or actual effects of the speech are central.[510] The IACmHR has held that Article 13 (5) requires proof of (a) the clear intention of promoting lawless violence or any other similar action against a protected group; and (b) the capacity to achieve this objective and create an actual risk of harm to such persons.[511] The ICCPR also requires the prohibition of incitement to discrimination, hostility or violence.[512] Incitement, in the context of this provision, has been defined as statements that create 'an imminent risk' of discrimination, hostility or violence.[513] This is categorised by the UN Special Rapporteur on the Freedom of Expression as perlocutionary acts of the speaker, seeking to provoke reactions in the audience.[514] In relation to incitement, the causality between speech and effect is generally considered closer than, for example, hatred. Although incitement is an inchoate crime and does not require that it is acted upon, the imminent risk or likelihood of that conduct occurring must be considered.[515]

Beyond this, there are divergences among adjudicatory bodies on whether to prohibit such acts as "advocacy" or "promotion" of hatred, which do not require an imminent risk of, for example, discrimination or violence. As mentioned, the Additional Protocol to the Budapest Convention defines hate speech as statements that advocate, promote or incite hatred, discrimination or violence. The ICERD also prohibits the mere act of disseminating racist ideas.[516] Furthermore, the CERD considers that racist hate speech does not involve solely explicit racist remarks but also indirect language disguising its objectives.[517] Such speech '. . .rejects the core human rights principles of human dignity and equality and seeks to degrade the standing of individuals and groups in the estimation of society'.[518] Meanwhile, the UN Strategy and Plan of Action on Hate Speech defines hate speech as speech 'that

[509] *Lewit v Austria* App no 4782/18 (ECtHR, 10 October 2019), para. 46.

[510] UN HRC, 'Report of the Special Rapporteur on Minority Issues, Rita Izsák' (2 January 2015), para. 54.

[511] IACHR, 'Annual Report of the Inter-American Commission on Human Right: Report of the Special Rapporteur for Freedom of Expression', Chapter III (Inter-American Legal Framework of the Right to Freedom of Expression) (30 December 2009) OEA/Ser.L/V/II. Doc. 51, para. 58.

[512] Art. 20 (2) of the ICCPR.

[513] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (7 September 2012), para. 44 (c).

[514] UNGA, 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (10 August 2011), para. 28.

[515] UN Rabat Plan of Action, para. 29. *See* also CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', paras. 13, 16.

[516] Art. 2 of the ICERD.

[517] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 7.

[518] ibid., para. 10.

attacks or uses pejorative or discriminatory language. . .'.[519] However, according to the UN Special Rapporteur on the Freedom of Expression, only the most egregious forms of hate speech should be prohibited, involving incitement to discrimination, hostility and violence. Expressions that raise concern in terms of tolerance and respect for others may be addressed through other measures, for example, education.[520] It should also be noted that as hate speech provisions involve speech denigrating a particular group, targeting an individual victim without seeking to incite others to take action against the group is not encompassed. Nevertheless, this may be prohibited by states in order to 'protect the rights and freedoms of others' in accordance with freedom of expression provisions, for example, in relation to hate crimes.[521] Additionally, negative stereotyping of a particular group, without involving hate speech, may be addressed as violations of the right to privacy—by affecting the reputation of the group—or as instances of discrimination, for example, under CEDAW.

As ample case law has come before the ECtHR on this issue, its approach will be further developed. As mentioned, the ECtHR has affirmed that while offending language is generally protected, hate speech is not. The ECtHR has not defined hate speech, but it clearly involves only those opinions that go beyond offensive, shocking or disturbing.[522] The Court has in its jurisprudence referred to the Recommendation of the Committee of Ministers on hate speech of 1997, which holds that the term "hate speech" is to be 'understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance. . .'.[523] The ECtHR does thus not in theory require incitement in order to exclude speech from protection, rather considering its harmful social effect in relation to the underlying values of the ECHR, such as significantly undermining democracy or political stability. Nonetheless, it appears from the combined case law of Article 17 and Article 10 that in practice a level of incitement to violence often is integral in order to categorise statements as "hate speech".[524]

---

[519] UN Strategy and Plan of Action on Hate Speech (2019), <https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf> Accessed 9 March 2022, p. 2. Similarly, UNESCO holds that hate speech provisions may extend to expressions that foster a climate of prejudice and intolerance, which in turn fuels discrimination and hostility. *See* Gagliardone et al. for UNESCO (2015), p. 11.

[520] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 24. *See* also UN HRC, 'Report of the Special Rapporteur on Minority Issues, Rita Izsák' (2 January 2015), para. 60.

[521] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 23.

[522] *Handyside v the United Kingdom* (1976) 1 EHRR 737, para. 49.

[523] *Gündüz v Turkey* (ECtHR), para. 22; *Féret v Belgium* (ECtHR), para. 44.

[524] For example, in *Sürek v Turkey (No. 1)* App no 26682/95 (ECtHR, 8 July 1999), para. 62, the Court considered that the appeal to bloody revenge and violence constituted hate speech. Meanwhile, in *Gündüz v Turkey* (ECtHR), para. 51, it held that merely defending Sharia, without inciting violence, could not be considered hate speech.

At the same time, the Court has made clear that incitement to hatred is not synonymous with calls for violence. Incitement is rather determined by the intent of the speaker to make someone else the instrument of his/her unlawful will,[525] and includes '[a]ttacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population'.[526] In *Vejdeland and Others v Sweden*, although the wording of the leaflets did not directly encourage individuals to commit hateful acts, they were 'serious and prejudicial allegations' and the state was justified in restricting such speech.[527] The Court in *Beizaras and Levickas v Lithuania* held that comments on a photograph of a same-sex couple kissing, including: 'it's not only the Jews that Hitler should have burned' and that 'faggots … [should be thrown] into the gas chamber' or have 'a free honeymoon trip to the crematorium', or have 'their heads smash[ed]' or be 'castrated' or be '[shot]', were not merely "obscenities", but hate speech.[528] It should be noted that the Additional Protocol to the Budapest Convention not only prohibits the dissemination of racist and xenophobic material and threats but also insults, although states may opt out of the latter.[529]

However, whether hate speech involves incitement to violence is relevant for the applicability of specific provisions in the ECHR. The ECtHR approaches hate speech from the viewpoint of either a legitimate restriction on the freedom of expression under Article 10 or—in cases involving Article 17—excluding such speech from protection under the ECHR. When assessing Article 17 cases, the Court at times categorises speech as "hate speech", but not consistently.[530] Similarly, states may prohibit speech under Article 10, regardless of whether it is labelled as hate speech by the Court. Article 10 (2) allows states to prohibit expressions, with states enjoying a wide margin of appreciation in relation to speech inciting violence or hate speech.[531] In reviewing the legitimacy of state restrictions of speech, the Court conducts a balancing test with the interference reviewed in its context, including the content of the statement, the circumstances in which the person made them and consideration of whether the interference was proportionate to a legitimate aim.[532]

Meanwhile, the application of Article 17, as opposed to Article 10 (2), entails that there is no balancing between individual and state interests involved in qualified rights. Speech is categorically excluded from Article 10 protection, thus involving a content-based restriction. The European Commission on Human Rights (ECommHR) has held that 'the general purpose of Article 17 is to prevent

[525] Tulkens (2012), p. 288.

[526] *Vejdeland and Others v Sweden* (ECtHR), para. 55, reiterating the *Féret v Belgium* case.

[527] ibid., para. 54.

[528] *Beizaras and Levickas v Lithuania* (ECtHR), para. 125.

[529] Art. 5 of the Additional Protocol to the Budapest Convention.

[530] *See* critique in Cannie and Voorhoof (2011), p. 77.

[531] *Karatas v Turkey* App no 23168/94 (ECtHR, 8 July 1999), para. 50.

[532] *Handyside v the United Kingdom* (ECtHR); *Jersild v Denmark* (ECtHR).

totalitarian groups from exploiting in their own interests the principles enunciated by the Convention'.[533] Activities aiming to encroach upon the essential values of democracy are included.[534] In *Glasenapp v Germany*, the Court held that the objectives of the protection of the rule of law and the democratic system takes precedence over the protection of the freedom of expression.[535] Accordingly, '. . .concrete expressions constituting hate speech, which may be insulting to particular individuals or groups, are not protected by Article 10 of the Convention'.[536] The duties and responsibilities attached to the exercise of the freedom of expression thus find a more solid foundation in Article 17. However, Article 17 has also in certain instances been applied indirectly as an interpretive guide when assessing the necessity of state interferences under Article 10 (2).[537]

In the case of *Lilliendahl v Iceland*, the Court clarified its approach to hate speech *vis-à-vis* these two articles.[538] It held that the concept involves two categories of speech: the first category comprises the gravest form of hate speech, which falls within the scope of Article 17 and the second category is 'less grave' and may be restricted by states under Article 10.[539] The gravest forms of hate speech involve calls for violence or other criminal acts, whereas the second category concerns speech insulting, holding up to ridicule or slandering a specific group. In the particular case, speech denigrating the LGBTQ community was 'serious, severely hurtful and prejudicial', as well as promoted 'intolerance and detestation of homosexual persons'. [540] Although constituting hate speech, it did not involve incitement to violence and was consequently assessed in relation to Article 10. Nevertheless, addressing hate speech solely in relation to Article 10 has been criticised. For example, in *Vejdeland and Others v Sweden*, Judge Yudivska argued that the majority should not have accepted the state aim of protecting the rights and reputation of others, that is, the case should not be viewed as a balancing exercise between the freedom of expression and protection of the targeted group. Rather, 'hate speech is destructive for democratic society as a whole', since such ideas gain credence, exacerbating discrimination and perhaps resulting in violence.[541] She further argued

[533] *W. P. and Others v Poland* App no 42264/98 (ECtHR, 2 September 2004).

[534] *Garaudy v France* App no 65831/01 (ECtHR, 24 June 2003).

[535] *Glasenapp v Germany* App no 9228/80 (ECmHR, 11 May 1984), para. 110.

[536] *Gündüz v Turkey* (ECtHR), para. 41, referring to the approach in *Jersild v Denmark* (ECtHR), para. 35. In *Glimmerveen and Hagenbeek v the Netherlands* App no 8348/78 & 8406/78 (Commission Decision, 11 October 1979), para. 196, the ECmHR held that the distribution of ideas encouraging discrimination is not protected by Art. 10, that is, it does not fall within the scope of the Article.

[537] *Garaudy v France* (ECtHR).

[538] *Carl Jóhann Lilliendahl v Iceland* (ECtHR), paras. 33–39.

[539] ibid., paras. 34–35.

[540] ibid., para. 38.

[541] ibid., para. 9.

that hate propaganda '. . .always inflicts harm, be it immediate or potential'.[542] As such, it should be excluded from protection under Article 10.

So far, the ECtHR has applied Article 17 in relation to racist,[543] anti-muslim,[544] and anti-semitic[545] speech. Totalitarian doctrines, such as the revival of extreme nationalism, have been categorised as incompatible with democracy and human rights. Since anti-semitic and racist ideas often are integral to totalitarian regimes, this appears to address the root causes of such political ideas.[546] At the same time, a tendency to stretch the material scope to any act incompatible with the Convention's underlying values of democracy and human rights has been noted.[547]

Moreover, hate speech is distinguished from speech criticising government policies.[548] However, the delineation between accepted political speech and incitement to racism and discrimination is not clearly drawn. As held by the Court in *Perinçek v Switzerland*, it is in the nature of political speech to be controversial and, frequently, virulent.[549] Statements may be acceptable if they are made in the context of a broader discussion of political ideas or public interests and there was in this case no intent to propagate discriminatory opinions. The fact that speech contains strongly worded statements does not *per se* affect the assessment, but rather the content of the arguments.[550] The intent of the statement and impact of the speech are also considered.[551] As argued in *Karatas v Turkey*, with a considerable impact of speech comes a greater risk of disrupting public order.[552] For example, in *Savva Terentyev v Russia*, a blog post containing offensive, insulting and virulent language against the police in Russia—likening them to pigs and filth and suggesting they be burned in an oven like in Auschwitz—was as a whole not considered incitement to hatred or violence.[553] The Court held that vulgar language may serve a stylistic purpose, and style is protected as an aspect of communication alongside the substance of ideas. [554] In this case, the statements were made in relation to a public discussion on the police suppressing political dissidents, that is, political speech. The mention of the Auschwitz ovens was considered a 'provocative metaphor' rather than a call for physical violence.[555]

---

[542] ibid., para. 11.

[543] *Glimmerveen and Hagenbeek v the Netherlands* (ECtHR).

[544] *Norwood v the United Kingdom* (ECtHR).

[545] *Ivanov v Russia* App no 35222/04 (ECtHR, 20 February 2007).

[546] Cannie and Voorhoof (2011), p. 63.

[547] ibid., p. 62.

[548] *Stomakhin v Russia* (ECtHR), para. 95.

[549] *Perinçek v Switzerland* App no 27510/08 (ECtHR, 15 October 2015), para. 231.

[550] *Stomakhin v Russia* (ECtHR), para. 114.

[551] *Perinçek v Switzerland* (ECtHR), paras. 232–233.

[552] *Karatas v Turkey* (ECtHR), para. 52.

[553] *Savva Terentyev v Russia* (ECtHR).

[554] ibid., para. 68.

[555] ibid., para. 72.

The categorisation of expressions as hate speech or speech excluded from protection of Article 10 is more or less contextual, which is also a common approach at the domestic level, given the connection between such speech and social institutions and norms.[556] The CERD and the UN Rabat Plan of Action also place an emphasis on context, that is, the social and political environment prevalent at the time the speech was made and disseminated, as well as the role of the speaker, the audience and the content of speech.[557] The UN Special Rapporteur on the Freedom of Expression has also held that factors for consideration are '. . .the existence of patterns of tension between religious or racial communities, discrimination against the targeted group, the tone and content of the speech, the person inciting hatred and the means of disseminating the expression of hate'.[558] Similarly, the ECtHR has contended that the margin of appreciation of states allows for a consideration of the harmful effects of speech from the viewpoint of historical, demographic and cultural contexts.[559] Statements may be considered hate speech in one context but not another and thus the content of speech cannot be analysed in the abstract. Furthermore, the intention is considered,[560] which affects the ability to convince, direct or incite the audience. The status of the speaker and the form and impact of the speech are also relevant.[561]

The identity of the audience is an additional factor. For example, speech can be perceived as humorous by some and as hate speech by others. In *Vejdeland and Others v Sweden*, the context was noted in relation to the impact of speech. The leaflets had been pressed upon young, sensitive and impressionable individuals.[562] However, although contextual in its approach, in practice the ECtHR has been rather categorical in relation to certain types of speech. An example is its treatment of Holocaust denial, which is generally excluded from protection.[563]

---

[556] *See*, for instance, *Erbakan v Turkey*, App no 59405/00 (ECtHR, 6 July 2006), para. 55; *Soulas and others v France* App no 15948/03 (ECtHR, 10 July 2008), para. 38; *Balsyte-Lideikiene v Lithuania* App no 72596/01 (ECtHR, 4 February 2009), para. 78; Brown (2015), p. 2.

[557] UN Rabat Plan of Action, para. 29, emphasises the consideration of contextual factors: (1) the content and form of speech; (2) the economic, social and political climate in the state; (3) the position or status of the speaker; (4) the reach of the speech, i.e. the audience and the means of transmission, e.g. whether on the Internet; (5) the objectives of the speech. *See* also CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 15.

[558] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (7 September 2012), para. 46.

[559] In *Soulas and Others v France* (ECtHR), para. 38, the Court evaluated the use of the terms "ethnic war" and "ritual rapes of European women". *See* also CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 15.

[560] *Gündüz v Turkey* (ECtHR), para. 51.

[561] *Sürek v Turkey (No. 1)*, paras. 59–62; *Karatas v Turkey* (ECtHR), para. 52 (poetry to a small audience).

[562] *Vejdeland and Others v Sweden* (ECtHR), para. 56.

[563] *Garaudy v France* (ECtHR). The UN Human Rights Committee has adopted a similar approach. *See Robert Faurisson v France*, Communication No. 550/1993, UN HRC, UN Doc. CCPR/C/58/D/550/1993 (19 July 1995).

As for obligations, the prohibition of disseminating materials inciting racial discrimination is explicit in the ICERD and the Additional Protocol to the Budapest Convention.[564] The European Council Framework Decision on combating certain forms and expressions of racism and xenophobia also calls for the criminalisation of hate speech.[565] In contrast, Article 17 of the ECHR merely *enables* states to take measures against acts or speech that may harm such ideals and does not impose an obligation. At the same time, according to Judge Tulkens, there is room to argue in favour of positive obligations to combat hate speech.[566] Such positive obligations have, for example, been affirmed in *Aksu v Turkey* regarding ethnic stereotyping, although in relation to Article 8[567] and it is implied in *Delfi v Estonia*, involving hate speech online.[568] In terms of the type of sanctions, such must be proportionate to the offence, with the ECtHR often preferring the use of civil remedies and fines, although not excluding criminalisation.[569] For example, in *Savva Terentyev v Russia*, the Court held that the criminal prosecution of speech is a serious measure, bearing in mind the availability of other means of intervention, and should be limited to cases of hate speech or incitement to violence.[570] This generates positive obligations to effectively investigate incidents that may constitute incitement to hatred and violence.[571] Although the ICERD requires that the dissemination of ideas based on racial superiority and acts of violence on the basis of such ideas are made offences punishable by domestic law, the Committee has also emphasised that criminalisation must be reserved for serious cases, while less serious cases can be addressed through other means.[572] This indicates that civil law remedies are also acceptable. The Committee has furthermore noted the efficacy of such measures as education and counter-speech, given that racist speech often stems from indoctrination or a lack of education.[573] Similarly, even though the ICCPR prohibits hate speech, it does not require proscription.[574]

[564] Art. 4 of the ICERD; Art. 3 and Art. 4 of the Additional Protocol to the Budapest Convention.

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

[566] Tulkens (2012), p. 284.

[567] *Aksu v Turkey* (ECtHR), para. 58.

[568] *Delfi v Estonia* (ECtHR), para. 110.

[569] The Court has, for example, stated that criminal sanctions for the glorification of terrorist acts are acceptable only if there is a risk that such attacks will be committed. *See Leroy v France* App no 36109/03 (ECtHR, 2 October 2008), para. 45. Cf *Yavus and Yaylali v Turkey* App no 12606/11 (ECtHR, 17 December 2013), para. 52.

[570] *Savva Terentyev v Russia* (ECtHR), para. 83. Also affirmed in *Beizaras and Levickas v Lithuania* (ECtHR), para. 111.

[571] *Beizaras and Levickas v Lithuania* (ECtHR), para. 129.

[572] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 12. *See* also UN Rabat Plan of Action, para. 34.

[573] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 30.

[574] Art. 20 (2) of the ICCPR.

4.4.3 *Online Hate Speech*

As the freedom of expression applies regardless of the medium, the approach to hate speech in international human rights law extends also to the Internet. This has been affirmed, for example, by the CEDAW Committee and the CERD, which have explicitly noted obligations to eliminate racist propaganda on the Internet.[575] In terms of online/offline coherence, according to the UN Special Rapporteur on the Freedom of Expression, the application of provisions on hate speech must be equivalent online, and not involve stricter penalties for individuals or excessively intrusive technological means of restricting speech.[576] Nevertheless, certain features of the Internet affect the transposition of the traditional approach to hate speech. This is evident in case law of the ECtHR involving online hate speech. As mentioned above, the ECtHR's assessment of the legitimacy of state restrictions of hate speech is contextual, in terms of its harmful effects. It takes into account not only the content of speech but also the forum in which speech is communicated, the audience and the extent of dissemination. This entails that the context of the Internet *per se* and the website in question has an impact on such factors, affecting the categorisation of statements as hate speech.

In *Delfi v Estonia*, the ECtHR at a general level held that the risk of harm caused by content on the Internet is higher than posed by the press, including hate speech, in view of its accessibility and capacity to communicate user-generated content at a global level, while being difficult to remove.[577] The CERD has also considered the heightened impact of statements on the Internet in assessments of the effect and reach of speech.[578] As discussed previously, the architecture of the Internet, its widespread impunity, coupled with sociological phenomena such as deindividuation and mob mentality, entail that existing beliefs tend to be affirmed, and racist and sexist ideologies radicalised. At the same time, the ECtHR has in cases concerning defamation held that expectations of decency must be lowered in view of social behaviour and norms on the Internet, as repugnant speech is more common and thus less harmful.[579] The sheer prevalence of hateful speech may lower the persuasive power of content and thus the risk of incitement to discrimination and violence. Additionally, the style of communication, such as vulgarity, is protected alongside

[575] CERD, 'Concluding Observations on Poland' (29 August 2019) UN Doc. CERD/C/POL/CO/22–24, para. 16 (b); CERD, 'Concluding Observations on the Combined Nineteenth to Twenty-First Periodic Reports of Sweden', Adopted by the Committee at its Eighty-Third session' (12–-30 August 2013) (30 August 2013) UN Doc. CERD/C/SWE/CO/19–21, para. 12; CEDAW, 'Concluding Observations on the Seventh Periodic Report of Finland' (10 March 2014) UN Doc. CEDAW/C/FIN/CO/7, para. 14.

[576] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 29.

[577] *Delfi v Estonia* (ECtHR), para. 133.

[578] CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 15.

[579] *Payam Tamiz v the United* Kingdom (ECtHR), para. 80; *Magyar Tartalomszolgaltatok Egyesûlete and Index.Hu Zrt v Hungary* (ECtHR), para. 77.

content, which encompasses much provocative speech online.[580] Consequently, while the prevalence of harmful speech is exacerbated on the Internet, standards of civility are reduced. This implies that individuals are not provided as extensive protection against hate speech in this sphere, which undermines an offline/online coherence from this perspective.

The Internet also challenges the consideration of cultural factors in the assessment of harm. Most material is accessible to all, both children and adults—unless the author has restricted access to a particular post or intermediary geolocation restrictions are employed—which may occasion widespread dissemination and a global, heterogenous audience. For example, while noting the necessity of considering the cultural context in which hateful statements are made, the ECtHR in *Perinçek v Switzerland* conceded that with the use of new electronic means of communication, messages cannot be regarded as purely local.[581] However, the Court did not accept the link between statements made in Switzerland and the treatment of Armenians in Turkey, that is, that there was a direct correlation. Although the remarks were not made on the Internet but during public events, the Court distinguished the case from *Vejdeland and Others v Sweden*—which involved a captured audience—and the distribution of leaflets reaching the entire population of a country.[582] The potential global reach of statements entails that legal assessments may need to become more categorical, that is, not relative to the particular cultural context.

In certain cases, the particular website in question and its readership have been viewed in isolation. In *Savva Terentyev v Russia*, virulent remarks were made on a blog.[583] The Court noted that the reach and potential impact of a statement on a blog with a small readership should not be approached in the same way as speech published on mainstream or highly visited webpages. Thus, '[i]t is. . .essential for the assessment of a potential influence of an online publication to determine the scope of its reach to the public'.[584] The Court criticised the domestic court for not considering whether the blog attracted many visitors and for not having tried to establish the actual number of users who had accessed the blog during the month in which the comment was available. Meanwhile, the Court took into account that the blog did not seem to attract much public attention, nor was the applicant a well-known blogger, popular user of social media, or a public or influential figure.[585] The UN Special Rapporteur on the Freedom of Expression has also held that a contextual approach must consider the means of dissemination: '[f]or example, a statement

[580] *Savva Terentyev v Russia* (ECtHR), para. 68.
[581] *Perinçek v Switzerland* (ECtHR), para. 246.
[582] ibid., para. 253.
[583] *Savva Terentyev v Russia* (ECtHR).
[584] ibid., para. 79.
[585] ibid., para. 81.

released by an individual to a small and restricted group of Facebook users does not carry the same weight as a statement published on a mainstream website. . .'.[586]

At the same time, in the more recent case of *Beizaras and Levickas v Lithuania*, the state argued that hate speech published on the applicant's Facebook page did not generate the same level of harm as on an Internet news portal, such as in *Delfi v Estonia*, which generally attracts more visibility and comments.[587] In response, the ECtHR argued that it did '. . .not find it unreasonable to hold that even the posting of a single hateful comment, let alone that such persons should be "killed", on the first applicant's Facebook page was sufficient to be taken seriously'.[588] It also rejected—categorically—that comments on Facebook 'are less dangerous' than those on Internet news portals, evident in this case, where an image posted on a Facebook page had 'gone viral', attracting more than 800 comments.[589] This, in a more insightful manner, considers modes of online communication.

In terms of liability for Internet intermediaries and media publishers, as noted in Sect. 3.4, the ECtHR has in instances of hate speech or incitement to violence in effect affirmed an obligation to actively monitor the comments sections of large news websites and remove unlawful comments without delay, given that such speech is "clearly unlawful" and does not require any linguistic or legal analysis.[590] It may involve algorithms or human moderation, or a combination of both. Such obligations have not been extended to Internet intermediaries as of yet, as they do not contribute their own content and thus do not control the material on the websites to the same extent. In these instances, a notice-and-takedown mechanism is generally encouraged. Similarly, in the proposed DSA of the EU, racist and xenophobic hate speech is categorised as "illegal" content and subject to obligations of notice-and-takedown for intermediaries. Additionally, the European Commission places obligations on intermediaries to remove or block "cyber incitement to violence or hatred" in the proposed directive encompassing this offence.[591] Nevertheless, as discussed previously, the view that such speech does not require complex linguistic or legal assessments must be countered. Beyond cultural and historic factors, the legal assessment considers the intent of the speaker, which also requires a contextual approach. This may be obscured online, given the lack of facial and tonal cues and new technological means of conveying meaning, such as multimodality. The developing approach by the ECtHR and the EU thus implies a turn toward more objective assessments of online speech.

[586] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue' (7 September 2012), para. 46.

[587] *Beizaras and Levickas v Lithuania* (ECtHR), para. 98.

[588] ibid., para. 127.

[589] ibid., para. 127.

[590] *Delfi v Estonia* (ECtHR).

[591] Art. 10 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

4.4.4 Sexist Hate Speech

4.4.4.1 Introduction

It is unclear what constitutes sexist hate speech as there is no widely accepted international definition. Similar to other gender-based online violations, different terminology is used, including "sexist hate speech",[592] "cybermisogony",[593] "cyber incitement to violence or hatred. . . directed at a group defined by sex or gender",[594] "hate speech against women"[595] and "gendered hate speech".[596] As to which forms of speech are encompassed, this has to a limited extent been developed by international organisations in soft law. For example, the Advisory Committee on Equal Opportunities for Women and Men of the European Commission defines sexist hate speech as 'threats of violence or public incitement to violence or hatred directed against a group of persons or a member of such a group defined on the basis of sex'.[597] The FEMM Committee of the European Parliament in turn describes it as 'expressions which spread, incite, promote or justify hatred based on sex'. [598] Meanwhile, the proposed EU directive on violence against women defines this offence as: '. . .intentional conduct or inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to sex or gender, by disseminating to the public material containing such incitement by means of information and communication technologies. . .'.[599] As such, certain general elements of hate speech are used, found in, for example, the ICERD, the ICCPR and in the jurisprudence of the ECtHR, generally applied to the protected ground of "sex" or "gender", that is, a gender-neutral approach. Nevertheless, the European Commission notes that incitement in this context may not necessarily reference the

[592] OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 16; CoE Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on preventing and combating sexism, 27 March 2019, p. 9.

[593] UN HRC, 'Right to privacy: Report of the Special Rapporteur on the right to privacy' (16 October 2019) UN Doc. A/HRC/40/63, para. 73.

[594] Art. 10 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

[595] Femm committee, European Parliament, 'Cyber violence and hate speech online against women' (2018), p. 45.

[596] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 68.

[597] Advisory Committee on Equal Opportunities for Women and Men, European Commission, 'Opinion on gender equality and the digital society in Europe: opportunities and risks' (2015), p. 3.

[598] Femm committee, European Parliament, 'Cyber violence and hate speech online against women' (2018), p. 18. *See,* also, OAS, 'Online Gender-Based Violence against Women and Girls: Guide of Basic Concepts, Digital Security Tools, and Response Strategies', p. 74.

[599] Art. 10 of the European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence'.

sex or gender of the victim, but the bias can be inferred from the content or context.[600] Such speech may be directed against a particular woman, on the basis of sex or gender, or women as a group. While focusing on incitement to violence, these definitions also encompass other acts. In a report by the CoE, anti-women websites, "slut-shaming", body-shaming, rape jokes, revenge porn, sexualised threats or offensive comments on appearance, sexuality and gender roles may be examples of sexist hate speech.[601]

With this approach, it is clear that sexist hate speech overlaps with other online gender-based offences, including sexual or other types of harassment.[602] These forms of speech also augment harmful gender stereotypes and thus require prevention, for example, according to Article 5 of CEDAW and privacy rights. However, as indicated, hate speech is considered a particularly grave offence, requiring a higher threshold of severity. It is limited to speech which incites or advocates discrimination, hatred or violence. It is thus reasonable to distinguish between sexist speech that embodies patriarchal beliefs and aims to influence people's ideas and values on the one hand, and speech which contains a coercive component on the other, that is, hate speech.[603] Accordingly, the CoE has made a distinction between sexism and sexist hate speech, in that the former is not necessarily characterised by hostility and thus encompasses, for example, sexist jokes, which reinforce gender stereotypes.[604] Certain soft law documents also include sexist hate speech in the concept of online harassment, or note the overlap between the two categories of offences.[605] However, as harassment and hate speech involve distinct legal assessments and state obligations—with hate speech requiring extensive measures, such as criminalisation and content moderation online—it is important to delineate such offences aligned with traditional concepts. Furthermore, categorising all instances of sexist speech as sexist hate speech may lead to a trivialisation of coercive and inciting speech against women.[606]

The aim of sexist hate speech, cyber-sexism and cyber gender harassment is understood as the humiliation or objectification of women, the undervaluing of their skills and opinions, the destruction of their reputation, instilling feelings of vulnerability and fear as well as the punishment of women for behaving in

[600] ibid., para. 22.

[601] CoE, 'Seminar Combating Sexist Hate Speech: Report', 10–12 February, EYC, Strasbourg, p. 16.

[602] Powell and Henry (2017), p. 169; Nussbaum (2010), p. 84.

[603] Richardson-Self (2018), p. 261; Sekowska-Kozlowska et al. (2022), p. 1.

[604] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 3.

[605] CoE (GREVIO), 'General Recommendation No. 1 on the digital dimension of violence against women', para. 39; European Institute for Gender Equality (EIGE), 'Cyber Violence against Women and Girls' (2017), p. 2.

[606] Sekowska-Kozlowska et al. (2022), p. 6.

non-stereotypical ways. [607] The latter is noticeable in that young women, women in the media or gaming, politicians, journalists and human rights defenders are targeted in particular, that is, frequently women in positions of power or visibility.[608] As argued by Danielle Citron, these are not isolated incidents of cyberbullying but rather systematic harassment of women, particularly against non-heterosexual, non-Caucasian women who transgress gender norms.[609] It is a manifestation 'within the overarching order of gender inequality and domination by men over women'.[610] Given the root causes of sexist hate speech, it precedes the Internet but is exacerbated online.[611] Its development is thus related to the context of hegemonic masculinity in all societies and the normalisation of sexualised and violent language, coupled with user anonymity, gaps in legislative frameworks and a lack of law enforcement.[612] The Internet is a place for cultural meaning-making. For example, phenomena such as memes contribute to norm creation.[613] As argued by Catharine MacKinnon: 'Words and images are how people are placed in hierarchies, how social stratification is made to seem inevitable and right'.[614] As such, it generates not only individual but also collective harm, such as fostering gender stereotypes and contributing to social exclusion, for example, with women withdrawing from the public sphere of the Internet.

Sexist hate speech is pervasive, particularly on the Internet. In a study by the CoE Youth Department in 2015, women were identified as one of the three main target groups of hate speech.[615] The prevalent impunity on the Internet has increased the level of hate speech in general, including sexist hate speech.[616] For example, 160 xenophobic websites were recorded in 1995 and approximately 4000 in 2002.

[607] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 23; CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 16; CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 4.

[608] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 7.

[609] Citron (2014), p. 14.

[610] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 4.

[611] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 15. *See* also UNHRC, '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), para. 20.

[612] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 16.

[613] Powell and Henry (2017), p. 98.

[614] MacKinnon (1993), p. 31.

[615] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 6. *See*, also, CoE Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on preventing and combating sexism, p. 8.

[616] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 3; European Commission, 'Proposal for a Directive of the European Parliament and of the Council on combating violence against women and domestic violence', para. 22.

Over half of these were hosted in the United States, by all accounts due to its liberal approach to the freedom of expression.[617] In 2009, the number was reportedly over 10,000.[618] Meanwhile, ideologies such as extreme nationalism and conservatism provide fertile ground for sexism.[619] As these ideologies increase, so does sexism. The emergence of online Incel communities has also been noted by the EU Commission as a factor exacerbating hostility towards women.[620] Although certain states have criminalised sexist hate speech, prosecution at the domestic level is generally limited as a result of ambiguous legislation, anonymous perpetrators and an unwillingness on the part of authorities to consider it a serious offence.[621] Given the exacerbation of such speech through the Internet, the following section will thus explore sexist speech in relation to the current international legal framework on hate speech, while arguing for the inclusion of "sex" or "gender" as protected grounds in hate speech provisions.

4.4.4.2 International Human Rights Law

The recent EU Directive on audiovisual media services obliges states to ensure that incitement to violence and hate speech on the basis of sex is removed on such online platforms, and the future EU directive on violence against women contains a similar provision on cyber incitement. Meanwhile, no explicit obligation for states to criminalise sexist hate speech exists in international human rights law treaties or have been developed through case law by regional human rights law courts or UN treaty bodies. Whereas the CERD has acknowledged gender-based aspects of racial discrimination—with the Committee calling for the recognition of the different life experiences of men and women—this rather affirms the necessity of an intersectional approach in relation to protected groups.[622]

Nevertheless, there is room to argue that obligations to prohibit sexist hate speech are included within the scope of existing norms, either correlating hate speech

[617] Reported in the General Assembly of the Council of Europe on the Draft Additional Protocol to the Convention on Cybercrime Concerning the Criminalisation of Acts of a Racist and Xenophobic Nature Committed through Computer Systems, Explanatory memorandum, Doc. No. 9538 (5 September 2002), para. 8.

[618] Daniels (2009), p. 5.

[619] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 17; UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 14.

[620] Communication from the Commission to the European Parliament and the Council: 'A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime', COM(2021) 777 final (9 December 2021), para. 3.2.3.

[621] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 9.

[622] CERD, 'General Recommendation No. 25: Gender Related Dimensions of Racial Discrimination', 56th session (20 March 2000), para. 1; CERD, 'General Recommendation No. 35: Combating Racist Hate Speech', para. 6.

regulation with general non-discrimination provisions or, more specifically, with CEDAW in order to include sex or gender as protected characteristics. As noted above, theories underlying prohibitions on hate speech, including linguistic theories, view such speech as forms of discrimination, that is, speech acts, or alternatively, as creating conditions conducive to discrimination. Approaching sexist hate speech as discrimination against women links it to both sex and gender, in that its victims are primarily women and the harassment gender-based.[623] The link between protected groups in the prohibition on discrimination to provisions on hate speech is in fact increasingly being made in international human rights law, which includes "sex". This correlation is reasonable, given that hate speech is frequently understood to include speech inciting discrimination and the values associated with regulation are mainly non-discrimination and human dignity.[624] The EU, IACmHR and the CoE have all made such a connection.[625] The UN Special Rapporteur on the Freedom of Expression has also advocated that protected groups in the prohibition on hate speech in the ICCPR should correspond to groups in the non-discrimination provision, thus including "sex".[626] A similar approach is observed in EU directives, encompassing the same groups as protected in the non-discrimination provision in the Charter on Fundamental Rights, which also comprises "sex".[627] As argued by the European Commission, victims of hate speech are targeted '. . .because of their immutable, unchangeable characteristics, or because of one that is at the core of their identity'.[628] Also the ECtHR has referenced the non-discrimination principle in Article 14 in cases involving hate speech, as an additional argument for restricting such speech.[629]

The CEDAW Committee has in several instances considered the specific issue of sexist hate speech, affirming that its proscription is an obligation emanating from the

[623] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 4.

[624] Art. 2 (1) of the Additional Protocol to the Budapest Convention; UN Rabat Plan of Action, para. 29 (a).

[625] The European Commission against Racism and Intolerance's (ECRI) General Policy Recommendation (GPR) No. 15 on combating hate speech, adopted on 8 December 2015, p. 3; IACmHR, 'Hate speech and Incitement to Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas' (31 December 2015), para. 18; CoE, Recommendation No. R(97)20 of the Committee of Ministers to Member States on "Hate Speech" Adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies), p. 107.

[626] UNGA, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' (9 October 2019), para. 9.

[627] *See*, for example, Art. 6 (1) (a) and Art. 9 (c) (ii) of the Audiovisual Media Services Directive (2018). *See* also Communication from the Commission to the European Parliament and the Council: 'A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime', COM(2021) 777 final (9 December 2021).

[628] Communication from the Commission to the European Parliament and the Council: 'A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime', COM(2021) 777 final (9 December 2021), para. 3.2.2.

[629] *Glimmerveen and Hagenbeek v the Netherlands* (ECtHR), para. 196.

prohibition on sex discrimination. In its Concluding Observation on Finland, the Committee raised concerns over the increase in online hate speech against women and girls, in particular against women belonging to ethnic minorities.[630] The Committee consequently called for the strengthening of measures to address hate speech against women and girls in the media, including on Internet discussion boards and on social media.[631] This issue was addressed in relation to "stereotypes and harmful practices", as an aspect of gender equality. In its Concluding Observation on Norway, the Committee also addressed the issue of sexist hate speech in relation to gender stereotypes and harmful practices.[632] The Norwegian penal code did not include gender as a protected ground in its hate speech provision, which was criticised by the Committee.[633] It called for the application of a special focus on the intersecting grounds of gender and race, ethnicity, religion and nationality in relation to hate speech.[634] Additionally, the Committee has recommended that Japan adopts legislation to prohibit and sanction sexist speech and propaganda advocating racial superiority or hatred against ethnic and minority women.[635] This was linked to criticism from the Committee concerning the frequent portrayal of women as sex objects and in stereotyped roles in the media.

Sexist hate speech is also part of the spectrum of violence against women. As mentioned above, violence against women is not defined solely as physical violence but includes verbal abuse. Furthermore, it can be argued that sexist hate speech increases the general risk of physical violence against women, by fostering an atmosphere of gendered subordination. Certain statements may be construed as direct incitement to violence against women. However, as noted, there are limited empirical studies on the causality between hate speech and violence. Nevertheless, while a direct link may be difficult to affirm, the gender stereotypical norms produced by such speech are linked to gender-based violence in international human rights law. This entails that a prohibition on sexist hate speech can also be derived from legal instruments on the combating of violence against women, such as the Istanbul Convention, Belém do Para Convention and the Maputo Protocol, in addition to general treaties prohibiting discrimination. For example, the UN Special Rapporteur on Violence against Women notes that states are obliged to protect women against harmful speech also in cases with no individual victims, for example, on online fora advocating violence against women.[636]

---

[630] CEDAW, 'Concluding Observations on the Seventh Periodic Report of Finland' (10 March 2014) UN Doc. CEDAW/C/FIN/CO/7, para. 14.

[631] ibid., para. 15 (c).

[632] CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017), para. 22.

[633] ibid., para. 22 (b).

[634] ibid., para. 23 (d).

[635] CEDAW, 'Concluding Observations on the Combined Seventh and Eighth Periodic Reports of Japan' (10 March 2016) UN Doc. CEDAW/C/JPN/CO/7–8, para. 21 (d).

[636] UNHRC, '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), para. 67.

Sexist hate speech also produces inequality *vis-à-vis* the protection of women's freedom of expression. As noted by the UN Special Rapporteur on the Freedom of Expression and the CoE, a failure to restrict sexist hate speech limits the guarantee of the freedom of expression, as it silences women and girls.[637] As noted, gender equality and the freedom of expression are mutually reinforcing and thus not in conflict *per se*. When female journalists are harassed, the public's right to be informed is impeded. It deprives society of a diversity of ideas, 'widening the democratic gap'.[638] Accordingly, 'allowing sexist hate speech to thrive with impunity does not constitute freedom of expression as it reduces plurality and diversity and is rather another manifestation of gender inequality'.[639]

Other soft law documents also indicate an increased interest in developing a regulatory framework on the matter. Primarily the CoE has taken steps to adopt regulation on sexist hate speech. General Policy Recommendation (GPR) N°15 is the first Council of Europe document that includes a definition of hate speech based on sex and gender.[640] Similarly, PACE Resolution 1751 (2010) on combating sexist stereotypes in the media calls on national parliaments to take measures to combat 'sexist stereotypes in the media by adopting legal measures to penalise sexist remarks or insults, incitement to gender-based hatred or violence and defamation of an individual or group of individuals on the grounds of their sex'.[641] Similarly, PACE Resolution 2144 (2017) and Committee of Minister recommendations call for the extension of the concept of hate speech to sexism. [642] The risk of harm of such speech is understood as involving an escalation or incitement to '. . .overtly offensive and threatening acts, including sexual abuse or violence. . .or potentially lethal action or. . . self-harm'.[643] The UN Strategy and Plan of Action on Hate Speech specifically includes "gender" as a protected ground.[644]

---

[637] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 70; CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 18.

[638] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 23.

[639] CoE, Prepared by the Gender Equality Unit, 'Background Note on Sexist Hate Speech' (1 February 2016), p. 5.

[640] ECRI General Policy Recommendation No. 15 on Combating Hate Speech (adopted on 8 December 2015), p. 3. *See* also CoE, 'Gender Equality Strategy 2014–2017' (February 2014), p. 10; CoE, 'Internet Governance – Council of Europe Strategy 2016–2019: Democracy, Human Rights and the Rule of Law in the Digital World' (Adopted at the 1252th Committee of Ministers' Deputies Meeting on 30 March 2016), para. 10 (d).

[641] PACE Resolution 1751 (2010) 'Combating sexist stereotypes in the media', para. 7.1.

[642] PACE Resolution 2144 (2017) 'Ending cyberdiscrimination and online hate', paras. 2 and 7.2.3; Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on preventing and combating sexism, p. 6.

[643] Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on preventing and combating sexism, p. 4.

[644] UN Strategy and Plan of Action on Hate Speech (2019), <https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf> Accessed 9 March 2022, p. 2.

Furthermore, beyond the correlation with protected groups in the non-discrimination principle, an evolutive treaty interpretation of provisions related to hate speech could allow for an expansion to "sex" and/or "gender" as protected grounds. Certain provisions, such as in the ICERD or the ICCPR, are exclusive in terms of enumerated groups, while others are open-ended and leave room for interpretation, such as in Article 13 (5) of the ACHR and Article 17 of the ECHR. Still, the UN Special Rapporteur on the Freedom of Expression has called for the recognition of gender-based hate speech under Article 20 (2) of the ICCPR, despite its limitations, in view of the gender equality clauses of the Convention.[645] States are also *allowed* to prohibit sexist hate speech in accordance with provisions on the freedom of expression, including under Article 10 (2) of the ECHR.

In terms of Article 17 of the ECHR, an analogy could be drawn to *Vejdeland and Others v Sweden*, where sexual orientation was considered an intimate aspect of a person's identity, allowing for state restrictions of the freedom of speech. The ECtHR in *Savva Terentyev v Russia* also noted a link between protected groups and historic prejudice and discrimination, and its approach to hate speech could thus extend to "sex", pertaining in particular to women. It should be borne in mind in this regard that the ECtHR has expressly rejected that "women" *per se* can be categorised as a vulnerable group, in contrast to certain sub-groups, such as Roma women.[646] However, this does not automatically entail that its jurisprudence on hate speech cannot extend to women. In consideration of sex being a characteristic fundamental to a person's identity as well as the historic, structural discrimination against women, it is a ground comparable to other protected identities. If "sex" or "gender" is encompassed as a protected ground, a similar delineation as in the jurisprudence of human rights law courts and treaty bodies of speech categorised as "hate speech" can be employed.

At the same time, without a direct correlation with non-discrimination provisions, rights interpretation may well exclude women. It appears that the general approach in international human rights law is to maintain exclusivity *vis-à-vis* this offence. For example, the UN Special Rapporteur on the Freedom of Expression has cautioned against the devaluing of the term hate speech, particularly in the digital age.[647] The reasons are multiple. Primarily, there is a perceived risk of excessively eroding the freedom of expression with an expansion of the concept. For example, the Philippines suggested the inclusion of the following provision during the drafting process of Article 5 CEDAW: 'Any advocacy of hatred for the feminine sex that constitutes incitement to discrimination against women shall be prohibited by law', which was opposed by the majority of states due to a fear that this would interfere with the

[645] UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 70.

[646] *See*, for example, *V. C. v Slovakia* App no 18968/07 (ECtHR, 8 November 2011), paras. 146, 178; *I. G., M. K. and R. H. v Slovakia* App no 15966/04 (ECtHR, 13 November 2012), para. 123.

[647] UN, 'Companies "failing" to address offline harm incited by online hate: UN expert' (21 October 2019) <https://news.un.org/en/story/2019/10/1049671> Accessed 9 March 2022.

freedom of speech and a perceived difficulty in separating illegal acts from innocent statements.[648] Also, the CoE CM Recommendation No. R (97) 20 on hate speech does not refer to sex, gender or gender identity, as it was considered necessary to avoid losing the focus of the text by covering all forms of intolerance (e.g. intolerance on grounds of sex, sexual orientation, age, handicap, etc).[649]

There may also be gendered ideological reasons underlying this approach. As argued by the Gender Equality Unit of the CoE, sexist hate speech is often seen as acceptable, harmless and less serious than other forms of hate speech, while it is one of the most widespread and systematic forms of hate.[650] Sexist speech is thus often trivialised and viewed as a personal matter, rather than one generating social harm. Because discrimination against women is currently ingrained in social institutions and structures, it may not be as overt as in relation to other groups, which may impede recognition.

One reason may be the pervasiveness of such speech, making it invisible and thus affecting the apparent utility of restrictions. For example, Alexander Brown argues that in delineating the categories of protected groups in relation to hate speech provisions, an option is to consider the balance of interests that maximises net satisfaction. The result may be different depending on the group in relation to risk factors, including whether the law will be overly broad or have a chilling effect on speech.[651] For example, he notes that gender-based insults are so common in ordinary language that to prohibit them would require the censoring of a substantial amount of speech.[652] As noted above, depending on the treaty in question, various forms of speech are encompassed, such as promotion, advocacy, dissemination or incitement. As noted, this would require a distinction between speech that constitutes incitement to gender-based violence and discrimination on the one hand and categorical statements of women as inferior, without elements of incitement, on the other. Since sexist speech is common in ordinary language, particular challenges arise in delineating the boundaries of unlawful expressions in this context. [653]

The Internet as a forum adds complexity, with the ECtHR considering the threshold higher for acceptable speech and the style of speech—such as vulgarity—protected. Given the prevalence of sexist and misogynistic speech online, it is conceivable that much speech will be categorised as merely offensive and not hate speech. This is also a factor in the potential delineation of liability for online platforms, whether involving the use of AI or human moderators. Assessments of sexist hate speech may accordingly be even more challenging than hate speech in general. Thus, in the balancing of rights, the prohibition of sexist hate speech may be

[648] Rehof (1993), p. 78.

[649] Explanatory Memorandum to CoE CM Recommendation No. R (97) 20 on hate speech, para. 22.

[650] CoE, 'Seminar Combating Sexist Hate Speech: Report', p. 6.

[651] Brown (2017), p. 33.

[652] ibid., p. 54. However, Brown does not consider this is a sufficient reason to refrain from regulation.

[653] ibid., p. 54.

considered overly restrictive of the freedom of expression, causing an excessive chilling of lawful speech. However, such a functional approach is not consistent with the equality principle and—particularly in view of the prevalence and exacerbation of sexist speech online—the development of regulation on sexist hate speech is necessary, also extending to intermediary liability. In fact, the terms of service of major social media platforms already cover gendered hate speech.[654] Given the high threshold in categorising speech as hate speech, an excessive erosion of the freedom of expression can be avoided. Additionally, the means in combating such forms of speech need not involve criminalisation but other forms of preventive measures. A graduated approach, considering different types of hateful rhetoric may thus be applied, as generally employed in international sources.

4.4.4.3 Conclusion

The prohibition of hate speech is contested, related to diverse theoretical approaches to the freedom of expression and its restrictions. This includes the viewpoint that speech cannot generate harm in the legal sense, or is less harmful than conduct, with the value of the freedom of expression considered superior. Accordingly, the social benefits of the freedom of expression outweigh restrictions. These alternative views are linked to the assessment of harm, with group-based harm difficult to prove empirically, whether involving the link between speech and violence or discrimination. A similar issue arises in relation to pornography, discussed in the next section. However, demands for affirmative empirical proof of the correlation between speech and such social ills as inequality entails that the means of confronting group-based harms become inadequate. As argued by Richard Posner: '[a] huge harm unlikely to materialize for several more years is not a lesser threat to the nation than a much smaller harm likely to materialize tomorrow'.[655] Nevertheless, a limited number of prohibitions of hate speech exist in international human rights law. These mainly take an exclusive approach in terms of protected groups and prohibited forms of speech, not explicitly recognising hate speech against women or on the basis of sex or gender. When women are considered, it is primarily in calls for an intersectional approach.

The reasons for the non-recognition of sexist hate speech may be multiple. As mentioned above, the provisions on hate speech at both the domestic and international level were developed in response to the events of World War II. Sexist hate speech is also pervasive and normalised and thus not recognised as particularly harmful. However, although there is no explicit obligation to prohibit sexist hate speech, it can tentatively be implied from CEDAW concluding observations, CoE soft law documents and developments in EU law, linking non-discrimination

[654] *See* in UNGA, 'Report of the UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (30 July 2021), para. 78.

[655] Posner (2006), p. 122.

provisions to the protection against hate speech. At a minimum, these indicate an increased willingness to develop the concept of sexist hate speech. This approach entails that in order to ensure equality and democracy, be it in relation to the freedom of expression (through encouraging a plurality of viewpoints), or to prevent gender-based violence (through prohibiting incitement to violence), measures to combat such speech must be adopted. The tremendous increase in hateful rhetoric online increases the calls for such a development. Although the Internet provides a platform for contesting harmful ideologies and developing more nuanced arguments to refute such ideas, the Internet as a gendered sphere—in relation to both accessibility and content—is in effect an unequal marketplace through which to confront such viewpoints. As the Internet is an integral aspect of the social lives of individuals and the dualism of the virtual and real worlds has diminished, the social norms propagated online are transferred to the non-virtual world, requiring the upholding of similar standards in both spheres.

What are then the benefits of addressing sexist speech as hate speech rather than individual instances of harassment? The offence of hate speech recognises the group-based and social impact of such forms of speech, which requires a focus on its root causes rather than solely addressing the symptoms. This categorisation would reduce the normalisation of such language by acknowledging its harmful effects, such as exacerbating discrimination. Hate speech is also categorically unprotected in terms of the freedom of expression, for example, in the ECHR. In certain treaties, positive obligations to prohibit hate speech are placed on states, rather than allowing for a balancing of interests, for example, as in relation to defamation. The ECtHR has even affirmed a positive obligation for states to ensure secondary liability for online news portals in relation to hate speech or incitement to violence, in effect requiring active monitoring. This does not extend to other forms of online speech. Thus, whereas gender stereotyping may be categorised as "harmful" speech, hate speech is clearly "illegal". Although monitoring obligations do not involve Internet intermediaries, social media corporations are increasingly developing technical means of detecting hate speech. Meanwhile, when the EU directive encompassing incitement to violence against women enters into force, clearer obligations for intermediaries to remove or block content will be established in this region.

However, difficulties in delineating a prohibition on sexist hate speech arise which, in certain respects, are general and not specific to gender. This includes defining what forms of speech constitute the promotion, advocacy or incitement to hatred, violence or discrimination, which may be common and integrated in daily language. The application of provisions to the Internet is particularly challenging from this standpoint. A low-register tone of communication is common on the Internet, with the ECtHR affirming that the "style" of writing is a protected aspect of the freedom of expression, including "vulgarity". The ECtHR has explicitly held that a higher threshold should be set for what constitutes unacceptable speech on the Internet. Under the cloak of anonymity, individuals are free to exaggerate, explore socially unacceptable fantasies, threaten and promote hatred with limited

repercussions. As misogyny is pervasive online, it means that harmful speech against women is allowed relative impunity, providing a breeding ground for sexist social norms.

The assessment of whether speech is categorised as hate speech is also contextual. It considers the historic, demographic and cultural circumstances of a particular state in assessments of the harm of speech. How this applies to a global audience is unclear. In practice, the ECtHR has indicated that it considers the particular website in question, mainly from the viewpoint of assessing the size of the audience and thus the impact of speech. Nevertheless, the Internet makes most websites void of context, commonly available to readers in a variety of states and cultural environments. In relation to online hate speech, the ECtHR has asserted that the legal and linguistic categorisation is not as complex as in relation to other speech-based offences, which indicates that cultural factors informing harm may be less important in this forum. Although online media publishers may be under an obligation to monitor comments, the increased use of algorithms to lexically detect hate speech also indicates a turn toward objective assessments. Although this may override subjective and gendered assumptions of what is harmful speech, there is also a risk that, given the prevalence and normalisation of sexist speech, incentives to regulate will be reduced in order to avoid excessive encroachments on the freedom of expression. Accordingly, the context of the Internet heightens the need for a gender-sensitive interpretation of the concept of hate speech and intermediary liability, which in turn may affect the development of more advanced technological solutions to prevent or remove such speech.

4.5 Harmful Pornography

4.5.1 Introduction

Much of the feminist discourse on pornography developed during a dynamic period in the 1980s and early 1990s.[656] While the discussion on the benefits and harm of pornography is thus not new in legal and feminist theory, and although the arguments are still relevant, it requires application to the context of the Internet. Foremost, the Internet significantly affects the consumers of pornography, exponentially increasing accessibility, regardless of the age of the users. Linked to this, studies indicate that the Internet has also influenced the content of pornography, contributing to more violent and demeaning material.[657] The Internet as a medium additionally affects the suitability and venues for regulation. Thus, the particular forum of the

[656] Dworkin (1989), MacKinnon (1985), MacKinnon (1993) and Lederer (1980a).

[657] Vera-Gray et al. (2021); Hughes (2002), p. 129; Bridges et al. (2007); UN Broadband Commission for Digital Development, Working Group on Broadband and Gender, 'Cyber-Violence against Women and Girls: A World-Wide Wake-Up Call', p. 7. *See*, however, the opposite result in Shor and Seida (2019).

Internet entails that an analysis of state obligations *vis-à-vis* regulation of such material merits a revisiting of existing theoretical and legal frameworks.

The presence of pornography *per se* or, alternatively, certain forms of pornography on the Internet will be analysed in relation to the broader concept of gender equality, but also as potentially causal of specific gendered cyber violations. It is thus proposed that the prevalence of harmful pornography online contributes to the manifestation of gender stereotypes, which is considered the root cause of, for example, online harassment, the non-consensual dissemination of intimate images and sexist hate speech. The following questions will consequently be addressed: Is pornography protected by the freedom of expression in international human rights law? What are the possibilities and/or obligations for states to prohibit pornography? Does it pertain to specific types of pornographic material? And, finally, does the Internet as a medium inform the content of state obligations?

4.5.2 Theorising Harm

Legal, feminist and linguistic theories are useful in several regards in evaluating the suitability of restricting the distribution of pornography. First, such theories provide a framework for analysing whether pornography is considered speech within the meaning of the freedom of expression, or whether such material constitutes speech acts, such as subordination. In relation to this, the theories provide a basis for discussing the effects, and thus the ensuing harm, of pornographic material, primarily in relation to women. Furthermore, they consider whether, even if such material causes harm, there is an overriding social benefit of non-regulation, for example, bearing in mind the values of the freedom of expression. Again, the forum of the Internet has an impact on all these aspects.

Assessing the harm of pornography, that is, whether and how it is harmful, is the most contested element in restricting such material, affecting its categorisation as speech. Non-regulation—primarily advanced through the theory on the freedom of expression as conducive to individual autonomy—is based on the view of pornography as either (1) harmless or; (2) not sufficiently harmful to override the protection of speech. Meanwhile, regulation is premised on the individual or social harm being of such gravity as to require a restriction of the freedom of expression. The main approach at the domestic level is that pornography, or certain forms of it, causes harm to the moral fabric of society.[658] A morality-based standpoint considers the harm to be the corruption and depravation of the morals of viewers and—potentially—society, and aims to preserve traditional social structures pertaining to relationships and sexuality.[659] Regulation thus serves to protect certain values and ideals—such as purity, chastity and fidelity—for the public good, as opposed to

[658] MacKinnon (1993), p. 105.

[659] Evans (2006), p. 90.

addressing the issue from the standpoint of gender equality. For example, pornography in the US and Australia is regulated through laws on obscenity, embodying a morality-based approach which restricts material that is 'morally offensive'.[660]

Several problems arise through such an approach, from both a theoretical and a legal perspective, on the correlation between harm and the law. As noted above, according to Mill, states should not regulate conduct unless it causes harm, understood mainly as physical harm. That is, harm to morality is not sufficient to warrant prohibition.[661] Additionally, feminist scholars have considered morality an unstable and gendered compass, which may be interpreted in a manner detrimental to women.[662] Similarly, at the international level, morality-based regulations are considered particularly ambiguous as there is limited consensus on morality, especially with regard to sexuality. This, for example, entails that state parties to the ECHR enjoy a broad margin of appreciation in relation to issues of public morals.[663]

Meanwhile, the categorisation of harmful pornography as speech is *per se* contested from a feminist viewpoint. According to radical feminists, pornography *constitutes* discrimination or *contributes* to discrimination.[664] As noted above, certain linguistic theories have also categorised pornography not as speech in the traditional sense, but as speech acts.[665] While pornography has been construed as a 'vehicle for the expression of ideas'[666] through communicating views on sexuality,[667] from a linguistic standpoint, words cannot be viewed separately from their effects, as '[s]ocial inequality is substantially created and enforced – that is, done – through words and images.'[668] Accordingly, from the viewpoint of pornography as illocutionary speech, it is argued that women are socially constructed to be sexually subservient to men, which is gender discrimination *per se*, that is, not a mere representation of women being degraded.[669]

The general approach among feminist legal scholars in the 1980s was not the elimination of erotica *per se*, but material that denigrated women, although certain feminists argued that this encompassed all pornography.[670] Pornography was considered an eroticisation of women's pain, defining sex as an act of male domination and thus encouraging the subordination of women.[671] The harm of pornography was

[660] ibid.

[661] Mill (1859), p. 22.

[662] McGlynn and Rackley (2009), p. 259.

[663] *Handyside v the United Kingdom* (ECtHR), para. 48.

[664] Dworkin (1989); MacKinnon (1993), p. 29.

[665] Langton (1993), pp. 307–308; MacKinnon (1993), p. 30.

[666] MacKinnon (1993), p. 14.

[667] Brigman (1983).

[668] MacKinnon (1993), p. 13.

[669] In turn, this view has been challenged, also from a linguistic perspective. *See* Butler (1997), p. 13.

[670] West (1993), p. 240 (certain forms of pornography); MacKinnon and Dworkin (1985), p. 121 (pornography *per se*).

[671] Dworkin (1989), p. xxxiii; MacKinnon (1993), p. 10.

accordingly perceived as threefold: (1) participating women are physically harmed in the production of pornography; (2) specific women are harmed as a result of the effect on the attitude of certain men who have consumed pornography, for example, leading to instances of sexual violence; and (3) all women are harmed through the ideology encouraged by pornography, in that it reinforces the view of women as sexual objects.[672] As such, this formulation of harm purposefully removed it from the area of morality, which places more limited obligations on states to restrict the freedom of expression.

The issue of direct harm to women participating in pornography engenders a complex debate on female agency and vulnerability, linked to theories on autonomy and the appropriate role of the state as liberal or protectionist. Whereas feminist legal scholars in the 1980s relied on case studies and anecdotes as support of such harm,[673] broader empirical studies on the subject are inconclusive.[674] Arguably, the view on direct harm affirms sexist stereotypes of both men and women, based on traditional gender roles in relation to sexuality.[675] It presumes a "false consciousness" of women who consent and adopts an approach parallel to child pornography, that is, as sexual acts between unequal partners.[676] Accordingly, the sexual liberation of women and their ability to control their own sexual identities is essential in achieving gender equality, with pornography being one venue of expression.[677]

However, the core of feminist theories on pornography focuses on the social harm of such material.[678] This mainly relates to its subordinating effects and potential incitement to sexual violence. To a degree, this aspect of harm also necessitates a balancing between the consent of participants and broader goals of gender equality, understood primarily from a radical feminist perspective. On the one hand, it can be argued that women should be free to make decisions that appear to make them complicit in patriarchal norms, since denying women agency is more harmful. [679] On the other hand, it is understood that an individual can make an autonomous decision, for example, to participate in harmful pornography, with such material undermining the autonomy of women at a structural level. [680] This entails that even if there is no direct harm to participants, social harm may ensue. The social harm rationale is, for example, evident in the definition of "child pornography" in the Lanzarote

[672] Decew (1984), pp. 84–86 This is an approach similar to the theorised harm of child pornography: (1) harm to the children depicted; (2) harm to other children exposed to child pornography or who are sexually abused as a result of the offender's exposure to the material and; (3) harm to society as a whole. *See* Leary (2008), p. 9.

[673] MacKinnon (1993), p. 20; Lederer (1980b); MacKinnon and Dworkin (1997), p. 205.

[674] *See* overviews in Russell (1980); Cawston (2019), p. 631; Shor and Seida (2020).

[675] Searles (1994), p. 477.

[676] Taylor (1994), p. 59.

[677] Searles (1994), p. 482; Sullivan (1992), p. 39.

[678] MacKinnon (1993), p. 48.

[679] Schneider (1986), p. 221.

[680] Langton (2011), p. 437.

Convention. It obliges states to prohibit simulated but realistic images of minors involved in sexually explicit conduct, that is, where no child is directly harmed, given the encouragement of sexual objectification of children through such images.[681]

According to the latter approach, pornographic images—where women are frequently situated in a subordinate position—influence social norms on the sexual relationship between men and women, power hierarchies and the legitimisation of violence against women.[682] Arguably, pornography sexualises and preserves inequality by promoting women as inferior objects to be used and controlled for male sexual pleasure.[683] This is done in a context of widespread sexual harassment and violence. The correlation of pornography with sexual violence and gender inequality is thus presumed and a principal aspect. However, it should be noted that such an approach may not involve all forms of pornography. Regulating expressions of an activity which is not unlawful itself—that is, sex—is complex.[684] It would, at a minimum, entail that there is a legitimate aim to restrict pornography depicting, or likely resulting, in the causing of a certain level of physical harm, as well as simulated rape.

An additional social harm is the impact of pornographic material on women's enjoyment of the freedom of expression. From a radical feminist perspective, pornography transforms women into objects and undermines their ability to speak, especially to resist sexual abuse.[685] Whereas certain feminists consider the appropriate remedy to be counter speech,[686] it is according to Catharine MacKinnon '. . .unrealistic to suggest that more pornography is the solution to the power imbalance rampant in our patriarchal society'.[687] This is a similar argument as in relation to sexist hate speech—that sexual oppression silences women through subordination and that regulation is thus necessary to ensure equal participation in the marketplace of ideas and the democratic process. However, feminist critique against censorship also exists from the viewpoint of the freedom of expression, in that it can be used by men to censor female sexual expression and thus silence women.[688] Accordingly, it may lead to a slippery slope of suppressing speech, such as certain forms of art, literature, websites with information on sexual identity or

[681] Art. 20 (3) of the Lanzarote Convention. *See* also CoE, 'Explanatory Report on the Convention on Cybercrime', paras. 93 and 102.

[682] Barry (1979), p. 211; Dworkin (1989); West (2000), p. 206.

[683] MacKinnon (1995), p. 1959; Brownmiller (1975), p. 394. According to MacKinnon, '[t]he more pornography there is, the more it sets de facto community standards, conforming views of what is acceptable to what is arousing, even as the stimulus to arousal must be more and more violating to work.' *See* MacKinnon (1993), p. 88.

[684] McGlynn and Rackley (2009), p. 246.

[685] Langton (1993), p. 299; MacKinnon (1985), p. 36.

[686] Taylor (1994), p. 61.

[687] Daum (2009), p. 563.

[688] Hunter and Law (1988).

reproductive rights, and especially target sexual- or other minorities, as such information may be controversial according to majority norms. [689]

Like the dichotomy on agency/helplessness, the dichotomy on social harm—that is, the cause and effect of such material—is unlikely to be resolved. Although the social harm is presumed by most feminist scholars from a theoretical viewpoint, a causal relationship between pornography and, for example, the prevalence of sexual violence is difficult to verify empirically. As discussed in relation to sexist hate speech, there is less empirical evidence of social harm than of individual harm in general, as the assessment of the physical, psychological or economic consequences at a group-level may be inconclusive. Nevertheless, certain empirical studies indicate a correlation between pornography and social harm.[690] The CEDAW Committee has also noted the influence of pornography in sexual abuse cases, on the basis of domestic statistics.[691] However, the causality is often not empirical beyond anecdotal statements.[692] The findings of such studies on the topic are also contested.[693] Certain studies in fact indicate an inverse result, that is, that sexual assault rates have declined in states notwithstanding the proliferation of pornography.[694] Accordingly, assuming causation between "viewing" and "doing", that is, what is viewed will be done to women, must be approached with caution.[695] A depiction may or may not be interpreted by viewers as coercive. If empirical support is lacking, criminal law may

[689] White (2006), p. 134.

[690] An overview in Dekeseredy and Carlson (2020); Evans (2006), p. 122; Wright and Funk (2014). Certain studies on child pornography indicate that such images are used by offenders for sexual gratification to groom children to be sexually molested, for example, by demonstrating to children that adult-child relationships are acceptable, to decrease the inhibitions of potential victims and as manuals for how children should please the offender. *See*, generally, Leary (2008), p. 13. Murray (2016), p. 405.

[691] CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017), para. 22 (d).

[692] For example, Michelle Evans uses the testimonies of certain former pornographic actresses and abused women as evidence of non-consent in participation or the use of pornography as a textbook for abuse. While noting the issue of causality, Evans argues that it is mainly a way of discrediting the narrative of women. *See* Evans (2006), pp. 99, 122. Meanwhile, MacKinnon argues that the traditional approach to causality and harm does not apply to the group-based harm *vis-à-vis* women. S*ee* MacKinnon (1984), p. 338.

[693] For example, Ronald Dworkin holds that although pornography is 'often grotesquely offensive' to women, 'no reputable study has concluded that pornography is a significant cause of sexual crime'. *See* Dworkin (1996), pp. 218, 230. Similarly, Joel Feinberg disputes causality with the argument that perpetrators of sexual violence are most likely already predisposed to violence at the time of consuming pornography. Accordingly, pornography may act incidentally as a catalyst but is dependent on the underlying psyche of the perpetrator and cultural patriarchal values. *See* Feinberg (1985), p. 155. *See* also Bennett (2016), p. 493; Searles (1994), p. 480, who argue that pornographic images do not cause social problems but rather depict such.

[694] *See*, generally, Ferguson et al. (2022); Gilden (2016), p. 476; Maris (2013), p. 14. *See* a study on pornography and gender equality: Barron (1990).

[695] Keller (1993).

become a tool for the protection of public morals and, as noted above, morality is generally considered an inappropriate object of penal protection.[696]

Nevertheless, the lack of empirical evidence of a direct causality between speech and harm does not necessarily preclude regulation.[697] For example, the Canadian Supreme Court in *R v Butler*—concerning the prohibition of violent, degrading or dehumanising sex portrayed in pornography—noted that although a direct causality between pornography and gender inequality is difficult to affirm, it is 'reasonable to presume' that exposure to harmful pornography has a causal impact on attitudes and beliefs. Graphic representations of an act portrayed as desirable necessarily has some effect on patterns of behaviour.[698] Similarly, it is presumed in international human rights law, without direct empirical support, that certain types of gender stereotypes are harmful to gender equality and generate violence against women. That is, even if the causality between certain forms of pornography and gender-based violence cannot be affirmed empirically, the gender stereotypes manifested in such material are considered harmful *per se*.

Nonetheless, although feminist scholars in general have aimed to delink pornography from morality, the question arises as to why feminist theories focus on the detrimental effects of pornographic material in particular, if the central aim is to discourage harmful stereotypes as an aspect of gender discrimination. Other content in the media, advertising, music, and literature also defines women in stereotyped roles.[699] As the same effort is not taken to regulate, for example, websites encouraging cosmetic surgery or anorexia, it arguably indicates that the movement is based on a particular moral standpoint, mainly corresponding with dominance feminism.[700] However, part of the reason lies in the presumption that women are subjected to actual sexual abuse when participating in such images and videos, as opposed to, for example, stereotypical advertisements. A difference is also made in relation to social harm. Pornography has been in focus due to its perceived link to especially grave harm, such as sexual violence and sexual harassment. Although other forms of speech also promote sexist ideas and gender stereotypes, pornography often explicitly depicts female sexual objectification. Given the fundamental role of sexual coercion as a means of subordinating women, material that perpetuates such sexual norms is deemed especially harmful.[701] It has furthermore been argued that pornographic material, as opposed to other forms of expressions, is a particularly

[696] Zavrsnik (2010), p. 180.

[697] *See*, for example, Mikkola (2019), p. 42.

[698] *R v Butler* [1992] 1 S.C.R. 452 (Supreme Court of Canada).

[699] As argued by Joel Feinberg, '[t]here is no doubt that much pornography does portray women in subservient positions, but if that is defamatory to women in anything like the legal sense, then so are soap commercials on TV. So are many novels, even some good ones. That some groups are portrayed in unflattering roles has not hitherto been a ground for the censorship of fiction or advertising'. *See* Feinberg (1985), p. 148. Ronald Dworkin argues in a similar manner, proposing that other material may in fact have a more direct causal effect. *See* Dworkin (1997), p. 228.

[700] White (2006), p. 88.

[701] Schwartzman (1999), p. 38.

persuasive method of conveyance.[702] The distinction thus appears to be based on the purported level of harm which, in relation to pornography, is the encouragement of sexual violence.

Nevertheless, the focus on pornography is not an argument not to extend the discourse further, involving other types of material on the Internet also harmful to gender equality. The CEDAW Committee has, for example, noted the detrimental impact of a range of gender stereotypes, including women confined to the role of housewives and mothers, in addition to sexual objects.[703] Similarly, radical feminist scholars do not dispute that other types of content exacerbate sexist norms.[704] However, such content may best be addressed through such state measures as education, rather than prohibition. Furthermore, the approach to harmful pornographic or gender-subordinating speech should also be viewed broadly and, for example, include misogynistic video games and other content on the Internet.

4.5.3 Online Pornography

Does the medium in which pornography is distributed have an impact on the above-discussed theoretical discourses? The bulk of pornographic material is currently marketed and distributed on the Internet.[705] It has been argued that pornography on the Internet poses no new conceptual problems but rather practical problems of regulation,[706] and that the mainstreaming of pornography on the Internet renders feminist arguments on censorship outdated.[707] However, there are indications that the medium affects the content of pornographic material, in addition to its reach and accessibility, and it is thus not merely a technical issue. As such, its accessibility online affects both the degree of social harm in addition to the form and viability of regulation. Feminist theories are also still relevant in addressing such material from the viewpoint of gender equality, which informs obligations in a public forum such as the Internet.

Pornography is moving further into the private sphere. In order to access pornography pre-internet, individuals would either physically buy material or have it delivered, which reduced consumption incentives. In contrast, pornographic content on the Internet is widespread—beyond jurisdictional boundaries—easily accessible and often free of charge, allowing individuals to view the material in their home. Meanwhile, increased accessibility affirms the legitimacy of content.[708] The

[702] West (1993), p. 241.

[703] *See* an overview in Cusack (2013).

[704] For example, in advertisement. *See* Lambiase et al. (2017), p. 39.

[705] Chaney (2012), p. 816.

[706] MacKinnon (1995), p. 1967.

[707] Maris (2013), p. 8.

[708] MacKinnon (1995), p. 1959.

prevalence of harmful material itself, be it child abuse images or violent pornography, thus creates an impression of acceptability, by providing positive reinforcement for abusers or consumers.[709] In this sense, although the Internet is solely a tool for disseminating material, the ease of distribution has affected the *content* of pornography. Since most pornography is published on the Internet, generating an immense market, the increasing competition between websites has pushed producers to develop more violent and degrading material.[710] Nevertheless, as a result of the lack of cohesion in defining "violent pornography", as well as the methods and materials used in such studies, the results are highly varied.[711] What is clear is that online pornography challenges the moral boundaries of viewers. Material that would previously have been found unacceptable flourishes online, with demeaning and violent pornography not solely relegated to the dark web but prevalent also on the main pornography websites.[712]

Furthermore, the Internet blurs the line between consumers and producers. Anyone with a digital camera can post amateur pornography on the Internet at little expense. The Internet has thus freed such material from the exclusive control of media conglomerates. These tend to be run by Western, heterosexual men, selecting the type of pornographic material produced on the basis of profit.[713] The Internet allows amateurs access to the market, directly influencing its content, and thus offering a greater variety of material. At the same time, research indicates that amateur online pornography has a higher level of female objectification and subordination than professional pornography.[714] Furthermore, with anonymity comes a growth in the market and, arguably, increased risks to those participating in the production.[715] Women and children are at times coerced into the production of online pornography, and the genuine nature of consent by adult participants in these instances must be assessed in a gender-sensitive manner. For example, the making of online pornography as part of domestic violence has been observed.[716]

---

[709] UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. x; Leary (2008), p. 13.

[710] Hughes (2002), p. 131; Evans (2007), p. 11; Bridges et al. (2007). According to the CoE, albeit in relation to child pornography, images are increasingly obscene, perverse and brutal. *See* Council of Europe, 'Group of Specialists on the Impact of the use of New Information Technologies on Trafficking in Human Beings for the Purpose of Sexual Exploitation: Final Report' (16 September 2003) EG-S-NT (2002) 9 Rev., p. 80. *See* also CRC, 'Concluding Observations on the Second Periodic Report of the United States of America Submitted under Article 12 of the Optional Protocol to the Convention on the Sale of Children, Child Prostitution and Child Pornography, Adopted by the Committee at its sixty-second session (14 January–1 February 2013)' (2 July 2013) UN Doc. CRC/C/OPSC/USA/CO/, para. 27. Nevertheless, certain research indicates that adult pornography online is not increasingly violent, for example, Shor and Seida (2019).

[711] Marshall and Miller (2019).

[712] Vera-Gray et al. (2021).

[713] Daum (2009), p. 546.

[714] Klaassen and Peter (2015).

[715] Daum (2009), p. 548.

[716] Evans (2007), pp. 11, 50.

The use of mobile phones has also led to an ease in capturing and storing images of sexual abuse, which leads to a higher risk of spontaneous offending behaviour, as the need for planning is reduced.[717] Arguably, user-generated pornography has paved the way for so-called revenge pornography and sexting, through normalising the filming of private sexual activities, with or without consent. Many pornographic websites feature a revenge pornography sub-genre.[718] However, empirical studies on causality are limited also in this regard. It should also be noted that pornography, prostitution and human trafficking in certain instances overlap on ICTs, through interactive cybersex, images, videos, webcam sex and chat rooms with textual pornography.

It is also increasingly common for minors to voluntarily create and distribute self-produced intimate material, be it for profit or social acceptance, on webpages, social networking sites or as messages sent through apps, such as sexting.[719] In certain instances, such behaviour is non-consensual, with girls more commonly coerced into sexting than boys.[720] Furthermore, gendered norms entail that sexually active boys, as opposed to girls, are more frequently revered. Thus, '[s]exting is not a gender-neutral practice'.[721] One study indicates that approximately 88% of self-generated sexually explicit content on the Internet has been uploaded elsewhere without permission, which further highlights the importance of emphasising privacy rights beyond the protection of secrecy. [722] If participants are underage, the distribution of sexually explicit material is considered illegal at the international level.[723] This includes material related to adolescents above the age of consent but below the age of majority.[724]

Additionally, challenges arise in restricting the audience on the Internet. The freedom of expression protects speech that may be unsuitable for certain age groups. However, on the Internet, such material is largely accessible also to vulnerable groups, such as children. Surveys indicate that adolescents are increasingly exposed to unwanted pornographic material, as a result of the marketing strategies of pornography production companies, such as pop-up ads, adware and spyware that

[717] McCartan and McAllister (2012), p. 261.

[718] Patton (2015), p. 408.

[719] Leary (2008), p. 5.

[720] Ross et al. (2019).

[721] Ringrose et al. (2012), p. 7.

[722] Internet Watch Foundation (2012) Study of Self-Generated Sexually Explicit Images & Videos Featuring Young People Online <https://www.iwf.org.uk/media/j5wphscw/iwf_study_self_generated_content_online_011112.pdf> Accessed 21 March 2022, p. 5.

[723] Art. 3 (c) of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (2002); Art. 5 of the Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography.

[724] However, the Lanzarote Convention allows states to exclude consensual material from prohibition when the child has reached the age of consent, to be determined by the state. *See* Art. 20 (3) of the Lanzarote Convention.

secretly install software that directs unknowing users to pornography sites.[725] According to a study in the UK, 57% of 9 to 19-year-olds who accessed the Internet at least once a week had come in contact with pornography. 10% of these actively searched for pornography and the rest were inadvertently exposed.[726] The question thus arises whether technological difficulties to ensure that material is accessed solely by an appropriate audience warrant restrictions on the publication of the material in general. That is, is it possible for a state to prohibit lawful online material on the basis that it is harmful to children? This access-oriented approach will be addressed further in the following section.

4.5.4 International Human Rights Law

4.5.4.1 Introduction

Pornography is considered a form of speech in international human rights law and its production and distribution are thus not considered violations *per se*. The protection of the freedom of expression extends also to information and ideas that offend, shock or disturb.[727] The fact that pornography is protected as a form of expression is made clear in case law. For example, the ECtHR in the 2019 case of *Pryanishnikov v Russia* held that the refusal to grant Mr. Pryanishnikov a film reproduction license for pornographic videos for audiences over the age of 18—on the basis of criminal investigations involving the applicant as a witness—violated his freedom of expression, in terms of due process guarantees.[728] Certain forms of pornography are also protected as an aspect of the right to privacy. For example, in *A.D.T. v the United Kingdom*, the ECtHR held that privately videotaped homosexual intercourse, which did not involve physical harm and was not publicly distributed, was protected by Article 8.[729] In contrast, the public showing of pornography for profit is not encompassed in the right to privacy.[730] This entails that pornography, at an abstract level without consideration of its specific content, is protected within the freedom of expression, be it in print or videos. It does not mean that it cannot be restricted for the purpose of a legitimate aim. The ensuing harm may, for example, override the

[725] Mitchell et al. (2007).

[726] Livingstone and Bober for London School of Economics (2005), p. 3.

[727] *Handyside v the United Kingdom* (ECtHR), para. 49.

[728] *Pryanishnikov v Russia* App no 25047/05 (ECtHR, 10 September 2019). The Court considered that the domestic courts had failed to balance the impact on the applicant's freedom of expression with the protection of morals and the rights of others. Accordingly, the restriction on distributing 1500 erotic films, for which Mr. Pryanishnikov owned the copyright, had a substantial impact on the applicant.

[729] *A.D.T. v the United Kingdom* App no 35765/97 (ECtHR, 31 July 2000).

[730] *Reiss v Austria* App no 23953/94 (Commission Decision, 6 September 1995). The display of a gay pornographic videotape in a bar was not private conduct protected by Art. 8.

interest in protecting speech. Although it is categorised as speech, pornography is, according to most theories on the freedom of expression, considered of low value. The threshold for regulation is thus lower than for certain other types of speech.

From the perspective of the freedom of expression, pornography is not treated as a uniform category, but rather the particular content of a website, video or magazine is assessed. Certain *forms* of pornography may allow for or require regulation. This primarily concerns offences involving child abuse images, which are prohibited in several international human rights law treaties, including the Optional Protocol to the CRC on child pornography,[731] the Budapest Convention,[732] and the Lanzarote Convention.[733] Regulation is also, to a more limited extent, required of violent pornography.[734] Many states distinguish between pornographic, obscene and extreme material in their domestic legislation.[735] A call on states to establish such a classification system and to oblige companies to submit their work prior to distribution has also been made in CoE soft law documents.[736] At the international level, the difficulty lies in the fact that there is no global standard of what is considered obscene, in addition to general domestic variations on restrictions of the freedom of expression. If regulation is a possibility, or even an obligation for states, it requires a determination of which material is considered harmful, not just for children, but at a social level. As mentioned above, restrictions on pornography are often approached as a matter of domestic morality, with states commonly providing "public morals" as the aim for regulation, for example, in cases before the ECtHR. This generates a wide margin of appreciation for states. As such, the ECtHR relies on domestic definitions of obscenity rather than having developed its own standards.

4.5.4.2 Controlling Access and Distribution

Limited guidance on the regulation of adult pornographic material exists in international human rights law sources. The ECtHR has in a few cases assessed whether

[731] Art. 1 of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (2002).

[732] Art. 9 of Convention on Cybercrime of the Council of Europe (Budapest Convention) (ETS No. 185) 23 November 2001.

[733] Art. 20 and Art. 21 of the Lanzarote Convention.

[734] *See*, below, the section on pornography and equality.

[735] For example, in the UK, there is a legal distinction between erotica, pornographic content, obscene material and extreme material, generating different forms of regulation, with solely extreme material requiring prohibition. *See*, the Crown Prosecution Services: <https://www.cps.gov.uk/legal-guidance/obscene-publications> Accessed 21 March 2022.

[736] CoE Parliamentary Assembly Resolution 1835: Violent and extreme pornography (2011); The CoE Gender Equality Strategy 2018–2023, para. 9.1.3; CoE Recommendation No. R (89) 7 of the Committee of Ministers to member states concerning principles on the distribution of videograms having a violent, brutal or pornographic content adopted on 27 April 1989, para. 1.

state regulation of obscene material has excessively interfered with the freedom of expression. Since pornography produced and viewed in the private sphere is protected, the cases mainly involve public exhibitions of such material, and thus the necessity of restricting access. In most cases, the aim provided by states is the protection of morals, with the Court often concurring that a pressing social need exists and that the measure is proportionate in relation to the aim.[737] Included in public morals is offence to 'intimate personal convictions'.[738] For example, in *Müller and Others v Switzerland*, the Court held that the confiscation of obscene paintings in a gallery was reasonable in relation to the aim of protecting morals, with due consideration of the unrestricted access to the exhibition, as there was no age limit.[739] In contrast, the ECtHR in *Scherer v Switzerland* considered that a book and video-store which included a back section seating 12 people where gay pornography was shown, found through word of mouth by customers, did not warrant prosecution for displaying obscene material.[740] The store was not discernible from the street and it was unlikely that the back-room would be visited by people who were unaware of the subject matter of the film. Access was restricted to adults who paid for admission. There was thus '. . .no danger of adults being confronted with the film against or without their intention to see it' and minors did not have access.[741] As such, the protection of morals did not provide a sufficient reason for interference.

This can be contrasted with the case of *Hoare v the United Kingdom*, concerning an applicant with a business publishing and distributing video tapes of a pornographic nature via mail order.[742] The videos were advertised in a national newspaper and the tapes were distributed upon request. The applicant and his wife were convicted of possessing and publishing obscene articles for profit. The appropriate standard of review at the domestic level was objective, considering whether the material had a tendency to "deprave and corrupt". The domestic court noted that the videos displayed an '. . .utter dehumanisation of the people concerned, particularly the women, who are almost invariably on the receiving end of what was going on'.[743] Although the state provided the protection of morals under Article 10 (2) as the aim of the interference, the Commission modified the claim and held that the issue rather concerned the aim of protecting the "rights of others".

The applicant complained that the conviction constituted a violation of his freedom of expression in Article 10. Though conceding that the material was obscene, he considered the prosecution disproportionate. Arguably, the videos did not deprave or corrupt citizens in general as only members of the public who shared his interests responded to the adverts. The European Commission of Human Rights

[737] *W.and K. v Switzerland* App no 16564/90 (Commission Decision, 8 April 1991).

[738] *Wingrove v the United Kingdom* (1996) 24 EHRR 1, para. 58.

[739] *Müller and Others v Switzerland* (1988) 13 EHRR 212.

[740] *Scherer v Switzerland* App no 17116/90 (Commission Report, 14 January 1993).

[741] ibid., para. 62.

[742] *Hoare v the United Kingdom* App no 31211/96 (Commission Decision, 2 July 1997).

[743] ibid., section: the facts.

noted that where '. . .no adult is confronted unintentionally or against his will with filmed matter, there must be particularly compelling reasons to justify an interference'. [744] In this case, it was unlikely that the videos would be purchased accidentally. However, referring to its approach in a previous case, it held that '. . .once they have been distributed, they can, in practice, be copied, lent, rented, sold and viewed in different homes, thereby escaping any form of control by the authorities'.[745] Accordingly, the conviction was deemed proportionate to the aim pursued.

Furthermore, in *Wingrove v the United Kingdom*, the Court held that using a box with a warning as to the film's content '. . .would have had only limited efficiency given the varied forms of transmission of video works'.[746] It could thus not be guaranteed that solely buyers, and not also minors, would view the films as videos may escape the control of authorities. Meanwhile, the Court in *Kaos GL v Turkey* held that a seizure of all copies of a magazine containing a few pornographic images was not justified since less restrictive means could have been used, such as a ban on selling the material to persons under 18 years of age, using a special cover with a warning addressed to minors, or selling it via subscription only.[747] A distinction has thus been made in relation to the possibilities of controlling the material.

Perrin v the United Kingdom is the sole case at the international level involving pornographic material on the Internet. It concerned a criminal law conviction for the publication of a webpage containing obscene images. Mr. Perrin had published a website with images of men engaged in fellatio, covered in faeces (coprophilia), pictures which were also accessible to any Internet user through a preview page. In order to view more images, the Internet user had to subscribe to the website. The criminal conviction was deemed justified by the ECtHR since states are allowed to protect public morals and/or the rights of others, and any Internet user could access the preview page, including children.[748] The Court implied that the issue would not have arisen had Perrin restricted access, for example, through age verification software. The Court further argued that even though prosecution may only result in limited protection of vulnerable people, this was no reason for states to abandon the attempt to protect them and that '. . .harm can be caused at any time at which a person is confronted with the material'.[749]

Similar to *Hoare v the United Kingdom*, criminal prosecution was thus deemed proportionate. States maintain a broad margin of appreciation in relation to public morals, and the Court did not find it disproportionate for a state to resort to criminal prosecution, especially when the other measures '. . .have not been shown to be more effective'.[750] The Court took into account the fact that the speech was commercial

[744] ibid., section: the law.

[745] ibid., section: the law, referring to *Wingrove v the United Kingdom* (ECtHR), para. 63.

[746] *Wingrove v the United Kingdom* (ECtHR), para. 63.

[747] *Kaos GL V Turkey* App no 4982/07 (ECtHR, 22 November 2016).

[748] *Perrin v the United Kingdom* App no 5446/03 (ECtHR, 18 October 2005).

[749] ibid.

[750] ibid., section D. Proportionality.

and the applicant stood to gain financially from the webpage, in holding that a financial penalty would not be a sufficient deterrent.[751]

These cases affirm that the specific content of the pornographic material is decisive as to whether restrictions are legitimate, but as this is primarily evaluated in relation to the public morals of states, the margin of appreciation is broad. This in turn creates legal uncertainty. How "public morals" is understood from a legal perspective is unclear and the Court and Commission have not expounded upon the level of obscenity suitable for regulation. However, in addition to morals, the protection of the rights and freedoms of others was also provided as a legitimate aim in several cases, which does not automatically entail a broad margin of appreciation. This aim appears primarily to be raised in order to protect children from viewing pornography, which is a particularly relevant consideration on the Internet.

Given the general approach on the offline/online coherence of human rights law norms, the freedom of expression should not be restricted to a greater extent online solely on the basis of its potentially wide reach. According to the EU Commission, this entails that content available in other media should not be prohibited online in order to, for example, protect minors.[752] A distinction must thus be made between material that is clearly unlawful—such as child abuse images—and pornographic material that is not unlawful but to which children should not have unrestricted access. With regard to the former, states incur obligations to eliminate material by, for example, adopting criminal laws involving production, distribution and use. In relation to pornographic material that is not unlawful *per se*, the issue is one of restricting access, primarily in relation to children.

Whether children suffer harm from accessing adult pornography is contested.[753] However, although there is limited guidance on restricting pornographic material at the international level, there are specific obligations to protect children from harmful material online. The Lanzarote Convention places positive obligations on states to protect children on the Internet. The Convention provides that children must be protected from witnessing sexual abuse or sexual activities online, thus corruption.[754] Whereas the CRC does not explicitly oblige states to restrict harmful material,[755] the Committee on the Rights of the Child has expressed its concern in several concluding observations over the fact that pornographic and other harmful materials are accessible to children on the Internet.[756] Similarly, several soft law

[751] ibid.

[752] Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, 'Illegal and Harmful Content on the Internet' (1996) 1996COM(96) 487, p. 18.

[753] Nair (2007). *See*, for example, Martellozzo et al. (2020); Andrie et al. (2021).

[754] Art. 22 of the Lanzarote Convention.

[755] However, states are obliged to adopt guidelines for the protection of children from material injurious to their well-being. *See* Art. 17 (e) of the CRC.

[756] *See*, for instance, CRC, 'Concluding Observations on Monaco' (9 July 2001) UN Doc. CRC/C/15/Add.158, para. 28; CRC, 'Concluding Observations on Croatia' (12 January 2005) UN Doc. CRC/C/143, paras. 202, 203.

sources, including the UN Special Rapporteur on the Freedom of Expression, affirm that states have an obligation to protect children from material that 'undermine their dignity and development'.[757] The exposure to such images is deemed to affect their current or future sexual and emotional development.[758] However, the content of state obligations in such sources is unclear and frequently involve awareness-raising efforts and education.[759] For example, the Istanbul Convention obliges states, in cooperation with private actors, to promote skills among children, parents and educators on how to deal with online content of a sexual or violent nature.[760] Nevertheless, the CoE Parliamentary Assembly has called on Member States to introduce and enforce sanctions for the sale of pornographic material to minors.[761] Likewise, the EU Directive on audiovisual media services places obligations on states to ensure that access to harmful content is not made available in such a way that minors normally encounter it, including on video-sharing platforms online.[762] This involves material that impairs '. . .the physical, mental or moral development of minors. . .'.[763] The proposed DSA of the EU also places obligations on VLOPs to restrict access to harmful material.[764]

Meanwhile, the above reviewed cases of the ECtHR involve negative obligations for states, that is, an evaluation of the legality of state interference, and not positive

[757] UNCHR, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Mr. Frank La Rue' (20 April 2010) UN Doc A/HRC/14/23, para. 52. *See* also UNODC, 'Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children' (2015), p. 13. The Committee of Ministers has noted the inadequate protection of children and young people against harmful content as a human rights issue pertinent to the Internet. *See* CoE, 'Recommendation CM/Rec(2012)4 of the Committee of Ministers to Member States on the Protection of Human Rights with Regard to Social Networking Services' (Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies), para. 3.

[758] CoE Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to Human Rights for Internet Users (Adopted by the Committee of Ministers on 16 April 2014 at the 1197th meeting of the Ministers' Deputies), Explanatory Memorandum, para. 94: in relation to '. . .online pornography, the degrading and stereotyped portrayal of women, the portrayal and glorification of violence. . .solicitation for sexual abuse purposes. . .which are capable of adversely affecting the physical, emotional and psychological well-being of children.'

[759] CoE, 'Recommendation CM/Rec(2018)7 of the Committee of Ministers of the Council of Europe on Guidelines to respect, protect and fulfil the rights of the child in the digital environment' (Adopted by the Committee of Ministers on 4 July 2018 at the 1321st meeting of the Ministers' Deputies), para. 48; CoE Recommendation CM/Rec(2014)6 of the Committee of Ministers to member States on a Guide to Human Rights for Internet Users, Explanatory Memorandum, para. 94.

[760] Art. 17 of the Istanbul Convention.

[761] CoE Parliamentary Assembly Resolution 1835: Violent and extreme pornography (2011); para. 9.2.2.

[762] Art. 6 (a) of the Audiovisual Media Services Directive (2018). This may be regulated through co-regulation or self-regulation (Art. 4 (a) (1)).

[763] ibid., Art. 6 (a).

[764] *See* Sect. 3.4.3.

obligations to protect individuals from harmful material, in contrast to obligations to ensure that media publishers remove hate speech online. At the same time, certain elements in the cases provide room for arguing that also positive obligations for states arise in terms of restricting access. For example, the lack of governmental control of material was a cause for concern in *Hoare v the United Kingdom*, which is relevant to the online context, where much obscene material is unrestricted. The Commission distinguished between material accessible to a limited audience and material that may be accessed unintentionally or against a person's will. As mentioned previously, there is certainly a higher risk of unintentionally being confronted with pornographic material on the Internet, for example, through adware or spam. Similarly, the Court considered the freely accessible preview page in *Perrin v the United Kingdom* and the lack of age restrictions in viewing material in *Müller v Switzerland* problematic. Although a solution may be age verification software, the approach by the European Commission of Human Rights in *Hoare v the United Kingdom* went even further. While the videos were sold solely to adults, the risk that such material could reach children entailed that regulation was warranted. Since states accordingly incur obligations in this regard, the issue cannot be resolved by solely relying on the initiative of Internet users to install, for example, child protective filters.

Nevertheless, whereas the control of access is generally accepted from a human rights law perspective in the offline world, given the difficulties of verifying the identity and age of users online, restrictions on the Internet have a more severe impact on the freedom of expression, in terms of impeding access for adults to lawful material.[765] As noted in the case of *Muscio v Italy*—concerning the unintentional exposure to pornographic material through spam mail—the Internet as a forum not only complicates practical possibilities for states to regulate spam, more suitably restricted through user filters, but it also reduces reasonable expectations of privacy and thus lowers protection. How this translates into various technical solutions on the Internet is unclear. Access to websites may be restricted through different means. For instance, the EU Directive on audiovisual media services suggests age verification tools, encryption and parental controls but also allows for stricter measures, to be determined by states.[766] As noted above, labelling is also an option. Nevertheless, these mechanisms are generally considered ineffective and easily circumvented. As material is disseminated by non-state actors, be it private individuals or corporations, liability schemes to hold these accountable must also be established. However, generally, such websites are not controlled by domestic authorities and it is thus difficult to restrain the dissemination of material. Furthermore, regulating the venues of pornography upholds the approach that pornography consumed in the private sphere is encompassed by the right to privacy and restrictions are solely required when entering the public sphere. In this regard, the Internet is categorised as a public

[765] Volokh (1996), p. 380.

[766] This includes encryption and parental controls. *See*, para. 20 and Art. 6 (a) (1) of the Audiovisual Media Services Directive (2018).

sphere where pockets of privacy can be ensured through technological means of limiting access.

The focus on the accessibility of material assesses the harm in relation to protecting children's emotional and physical development and not the effects on society at large. Hence, general social harm—such as gender inequality—is not considered from such a perspective. The acknowledgment of such harm would *require* a prohibition of, at a minimum, violent and degrading material, regardless of the audience. In relation to the emerging distinction between illegal or harmful online material, it would also strengthen the argument for categorising certain forms of pornography as the former. It should be noted that Judge Pinto de Albuquerque in *Pryanishnikov v Russia* of the ECtHR criticised the majority precisely for failing to explore the increasing support for restricting pornography at the international level, particularly extreme pornography, in the balancing of interests. [767]

4.5.4.3 Pornography and Gender Equality

Analysing material solely from the viewpoint of the freedom of expression rarely involves the delineation of positive obligations to restrict speech. This is partly due to the unwillingness to categorise certain types of speech as harmful *per se*. In contrast, feminist legal scholars consider particular forms of pornographic material to be either discrimination or instruments of subordination, which in turn is linked to gender-based violence.

At the international level, such a link is occasionally made by the CEDAW Committee, placing positive obligations on states to restrict pornographic material harmful to gender equality. For example, in General Recommendation No. 19, the CEDAW Committee stated that: '[t]raditional attitudes by which women are regarded as subordinate to men. . .contribute to the propagation of pornography and the depiction and other commercial exploitation of women as sexual objects, rather than as individuals. This in turn contributes to gender-based violence'. [768] On the basis of such a link, the CEDAW Committee has approved of the adoption of legislation in states aimed at curtailing harmful pornographic material. [769] Meanwhile, in its Concluding Observation on Sweden in 2016, the Committee expressed its concern over the prevalence of pornography and the 'sexualisation of the public sphere', including stereotyped and sexualized images of women in the media and

[767] Accordingly, '[p]ornography frequently desensitises the consumer to sexual aggression, normalises sexual assault and promotes a rape culture, which impacts seriously on gender equality'. *See Pryanishnikov v Russia* (ECtHR), para. 29. A correlation was also drawn to gender-based violence, most pressing in relation to pornography depicting sexual violence, with Judge Pinto de Albuquerque holding that '[e]xtreme pornography contributes, directly and indirectly, to violence against women'. *See* ibid., Concurring opinion of Judge Pinto de Albuquerque, paras. 30–31.

[768] CEDAW, 'General Recommendation No. 19 on Violence Against Women', paras. 11 and 12.

[769] CEDAW, 'Concluding Observation on New Zealand' (12 April 1994) UN doc. A/49/38, para. 658 (c).

advertising sector, noting that it '. . .may exacerbate sexual harassment and gender-based violence against women and girls'.[770] It recommended the state to engage with companies and impose stricter regulations in order to enhance positive and non-stereotypical portrayals of women in the media.[771] In relation to Japan, it expressed concern over pornography, video games and animation that promote sexual violence against women and girls, calling on the state to implement effective legal measures and to monitor the regulation of the production and distribution of pornographic material that '. . .exacerbate discriminatory gender stereotypes and reinforce sexual violence against women and girls'.[772] Finally, the Committee in a concluding observation on Norway noted the influence of pornography in abuse cases, with a link made on the basis of national statistics, indicating an increase of 60% of reported rape cases in 2016, where the alleged perpetrator was a child.[773]

Furthermore, the UN Human Rights Committee in its General Comment No. 28 has urged states to inform the Committee on laws or policies that may impede women in exercising their freedom of expression on an equal basis and '[a]s the publication and dissemination of obscene and pornographic material which portrays women and girls as objects of violence or degrading or inhuman treatment is likely to promote these kinds of treatment of women and girls', the information must include measures to restrict the publication or dissemination of such material.[774] Similarly, the Beijing Declaration and Platform for Action provides: '[i]mages in the media of violence against women, in particular those that depict rape or sexual slavery as well as the use of women and girls as sex objects, including pornography, are factors contributing to the continued prevalence of such violence, adversely influencing the community at large, in particular children and young people'.[775] Furthermore, the CoE has in numerous documents affirmed a link between violent pornography and gender-based violence, noting '. . . .the desensitisation resulting from continued exposure or addiction to pornography, and against a process of normalisation in which moral coercion and physical violence may be considered as acceptable'.[776]

[770] CEDAW, 'Concluding Observations on the Combined Eighth and Ninth Periodic Reports of Sweden' (10 March 2016) UN Doc. CEDAW/C/SWE/CO/8–9, para. 24 (c).

[771] ibid., para. 25 (b).

[772] ibid., paras. 20 (c) and para. 21 (b).

[773] CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017) UN Doc. CEDAW/C/NOR/CO/9, para. 22 (d). This was addressed under the heading of "stereotypes and harmful practices".

[774] UN HRC, 'General Comment No. 28: Article 3 (The Equality of Rights Between Men and Women)' (20 March 2000) UN Doc. CCPR/C/21/Rev.1/Add.10, para. 22.

[775] Beijing Declaration and Platform for Action (Platform for Action), para. 118. Meanwhile, Art. 13 (m) of the Maputo Protocol obliges states to '. . .take effective legislative and administrative measures to prevent the exploitation and abuse of women in advertising and pornography'. This is not a prohibition on pornography *per se*, but rather involves coerced participation, that is, the direct harm principle. *See* the interpretation in ACmHPR, 'The Guidelines on Combating Sexual Violence and its Consequences in Africa' (2017), para. 3.1 b.

[776] CoE Parliamentary Assembly Resolution 1835: Violent and extreme pornography (2011), para. 6; The CoE Gender Equality Strategy 2018–2023, para. 40. CoE Recommendation No. R (89) 7 of

Moreover, the UN Special Rapporteur on Violence against Women considers that '. . . .most pornography represents a form of violence against women and that. . .evidence shows that it is directly causative of further violence against women', with the Internet making hard-core pornography mainstream.[777] It should, however, be noted that while the latter report proposes that research demonstrates that watching pornography increases men's propensity for violence as well as causes direct harm to participants, limited sources are referenced.[778]

These international bodies thus indicate state obligations to restrict harmful material, both through legislation and other measures, such as education on gender stereotypes and new technologies. The basis for this standpoint is the acceptance of a causality between the prevalence of pornography and gender-based harm against women, including sexual harassment and sexual violence—that is, gender inequality—primarily through the production of harmful gender stereotypes, such as rape myths. The main approach is that such material is not discriminatory *per se*, but rather contributes to an environment conducive to subordination, primarily in the form of sexual violence. Causality between such material and social harm is principally presumed on the basis of a theoretical framework, although empirical studies are also mentioned in certain instances.[779] In contrast, the direct harm approach has generally not been accepted at the international level in relation to adult women, solely children. [780]

Does the approach to harm as that of inequality pertain to pornography *per se* or certain types of pornographic material? The CEDAW Committee has in certain instances expressed concern over the prevalence of pornography in general but has mainly focused on harmful stereotyped depictions of women. Similarly, the other sources address the issue generally from the viewpoint of gender equality but

---

the Committee of Ministers to member states concerning principles on the distribution of videograms having a violent, brutal or pornographic content adopted on 27 April 1989, para. 6, encourages 'member states [to] consider if the application of their criminal law concerning videograms is effective in dealing with the problem of videograms having a violent, brutal or pornographic content. . .'.

[777] UN Commission on Human Rights, 'Preliminary Report Submitted by the Special Rapporteur on Violence against Women, its Causes and Consequences, Ms. Radhika Coomaraswamy, in Accordance with Commission on Human Rights Resolution 1994/45' (22 November 1994), para. 240.

[778] ibid., para. 237. The UN Special Rapporteur based her assertion on both empirical studies as well as the theoretical framework developed by Andrea Dworkin and Catharine MacKinnon.

[779] The CEDAW Committee has also accepted domestic statistics as a basis. *See*, for example, CEDAW, 'Concluding Observations on the Ninth Periodic Report of Norway' (22 November 2017), para. 22 (d). It also recommended that Sweden conduct a study on the possible impact of over-sexualized representations of women and girls in the media and the correlation between the prevalence of pornography and gender-based violence against women. *See* CEDAW, 'Concluding Observations on the Combined Eighth and Ninth Periodic Reports of Sweden' (10 March 2016), para. 25 (c).

[780] In contrast, Art. 20 of the Budapest Convention encompasses both the direct harm and indirect harm principle.

particularly denounce pornography that contains stereotyped portrayals of women, where women are objects of degradation. Whereas these sources thus indicate certain elements of harmful pornography—including the depiction of sexual or other forms of violence and subordination or other harmful gender stereotypes—further concretisation is necessary with a view to delineating state obligations.

Additionally, the categorisation of such content as harmful to society in general, and not solely children, requires more extensive obligations for states, beyond restricting access. However, what such obligations entail, beyond general measures of education, is similarly unclear and must be specified in the context of the Internet. Whereas it may include regulation restricting distribution of such materials, in terms of intermediary liability such content would under the current regime still be categorised as "harmful", placing limited obligations on intermediaries. This solely extends to due diligence considerations of VLOPs under the DSA, unless categorised as illegal at the domestic level. More extensive and concrete obligations for states to regulate both publishers and intermediaries in relation to violent and objectifying pornography is thus warranted.

4.5.5 Conclusion

Pornography is a contested area in domestic and international law, as a result of cultural variations on morals, sexuality and gender roles. It also brings to the fore fundamental differences on the values of the freedom of expression and theoretical approaches to harm. Whereas liberalism is premised on an unrestrained freedom of expression and allows state restraint solely in cases of empirically-based individual and—preferably—physical harm, the feminist discourse on pornography provides a theoretical basis for linking pornography to such social harms as gender inequality, which would generate positive obligations to restrict material. Even if conceding to the harmful effects of pornography, the liberalist approach considers it less harmful than restraints on fundamental rights, that is, it is outweighed by the importance of the freedom of speech.[781] In contrast, Catharine MacKinnon has argued: 'if a woman is subjugated, why should it matter that the work has other value?'[782]

If it is accepted that certain forms of pornography are harmful, delineations of the type of material that warrants regulation must be made. This, again, correlates with the presumed harm. If the harm is understood as the production and exacerbation of gender stereotypes, regardless of whether depictions of violence are involved, much pornographic material would be subject to restraint. If restrictions solely involve pornography portraying violence against women, the scope of state obligations would be reduced. The form of state obligations is also contested, for example, whether it involves censorship, criminalisation or broader efforts of education on

[781] Strossen (1996), p. 478.

[782] MacKinnon (1985), p. 21.

harmful gender stereotypes. For instance, if involving pornographic material that does not contain violence, restricting access for children or unintentional viewers may be an adequate response. Furthermore, pornography exists in a culture of pervasive sexism, of which it is a symptom. It could thus be argued that rather than criminalising the material, the harmful root causes—such as cultural norms on sexuality—should be addressed.[783] At the same time, regardless of whether harmful pornographic material is a cause or effect, it contributes to the pervasiveness of harmful sexist norms and thus warrants regulation.

It appears that there are no obligations at the international level to prohibit pornographic material *per se*, as it constitutes a form of speech. In contrast, child abuse images are prohibited since they are evidence of sexual abuse of a child, while also generating social harm. Whereas the regulation of pornography, according to the ECtHR, falls within the scope of the margin of appreciation of states, generally on the basis of morals, certain other sources have affirmed an obligation to restrict pornography, particularly obscene material. The CEDAW Committee and the UN HRC have approached such forms of pornography from the standpoint of its impact on gender equality, by contributing to harmful gender stereotypes and exacerbating the prevalence of gender-based violence. Although the causality between pornography and such group-based effects as increased levels of gendered violence is contested, it is tentatively made by these bodies. However, even if such causality is difficult to verify empirically, obligations to eliminate gender stereotypes do not require evidence of a particular effect. Consequently, the specific content, centred on women as objects in stereotypical roles, with eroticisation of non-consent, is considered harmful *per se*. This is also the view in various soft law documents by the CoE and the UN Special Rapporteur on Violence against Women. With such an approach, harmful stereotypes in material other than pornography may be addressed in a similar fashion.

As for suitable state measures to control material accessible on the Internet, unlawful pornography must be prohibited and removed, whereas general access to other types of pornographic material must be restricted, for example, to protect children. Where the context of the Internet has been considered in international human rights law, much focus has been placed on controlling access. Stricter obligations consequently ensue in cases of pornography categorised as unlawful (e.g. child abuse images), whereas technical solutions such as age verification systems or filters may be sufficient in other cases. For example, in terms of the protection of individuals against the unintentional viewing of pornography online, there is a tentative indication that responsibility is mainly placed on Internet users to install filters, due to the technical difficulties in regulating spam and malware. It is thus of the utmost importance to determine and expand on which types of pornographic material—beyond child abuse images—are considered unlawful. In terms of assessing the effect of pornography on the Internet beyond access, the fact that such material is more extreme and widespread online has not been specifically noted in

[783] Maris (2013), p. 15.

international sources. The easy access entails that harmful stereotypes of women and sexual violence are made global, mainstream and thus portrayed as acceptable. This warrants effective state obligations in restricting such material, be it through criminalisation or civil legislation, as well as liability schemes for Internet intermediaries. Developing an international human rights law approach that considers the well-established feminist framework on pornography, in light of the particular architecture and social norms on the Internet, is thus necessary.

References

Andrie EK et al (2021) Adolescents' online pornography exposure and its relationship to sociodemographic and psychopathological correlates: a cross-sectional study in six European countries. Children 8:1–16

Angelopoulos C, Smet S (2016) Notice-and-fair-balance: how to reach a compromise between fundamental rights in European intermediary liability. J Media Law 8:266–301

Barak A (2005) Sexual harassment on the internet. Soc Sci Comput Rev 23:77–92

Bardall G (2017) The role of information and communication technologies in facilitating and resisting gendered forms of political violence. In: Segrave M, Vitis L (eds) Gender, technology and violence. Routledge, Oxford, pp 100–117

Barendt E (2009) Balancing freedom of expression and privacy: the jurisprudence of the Strasbourg Court. J Media Law 1:49–72

Barron L (1990) Pornography and gender equality: an empirical analysis. J Sex Res 27:363–380

Barry K (1979) Female sexual slavery. New York University Press, New York

Bell MC (2021) John Stuart Mill's harm principle and free speech: expanding the notion of harm. Utilitas 33:162–179

Bennett JT (2016) The harm in hate speech: a critique of the empirical and legal bases of hate speech regulation. Hastings Const Law Q 43:445–536

Brennan F (2009) Legislating against internet race hate. Inf Commun Technol Law 18:123–153

Brenner A (2013) Resisting simple dichotomies: critiquing narratives of victims, perpetrators, and harm in feminist theories of rape. Harv J Law Gend 36:503–568

Bridges A et al (2007) Aggression and sexual behavior in best-selling pornography: a content analysis update. Violence Against Women 16:1065–1085

Brigman W (1983) Pornography as political expression. J Pop Cult 17:129–134

Brown A (2015) Hate speech law: a philosophical examination. Routledge, New York

Brown A (2017) The who question in the hate speech debate: part 2: functional and democratic approaches. Can J Law Jurisprud 30:23–55

Brownmiller S (1975) Against our will: men, women, and rape. Simon and Schuster, New York

Burgess MCR et al (2011) Playing with prejudice: the prevalence and consequences of racial stereotypes in video games. Media Psychol 14:289–311

Butler J (1997) Excitable speech: a politics of the performative. Routledge, New York

Cannie H, Voorhoof D (2011) The abuse clause and freedom of expression in the European Human Rights Convention: an added value for democracy and human rights protection? Neth Q Hum Rights 29:54–83

Cawston A (2019) The feminist case against pornography: a review and re-evaluation. Inquiry 62: 624–658

Chaffin S (2008) The new playground bullies of cyberspace: online peer sexual harassment. Howard Law J 51:773–818

Chaney N (2012) Cybersex: protecting sexual content in the digital age. J Marshall Rev Intellect Prop Law 11:815–840

Chinkin C (2004) Sexual harassment: an international human rights law perspective. In: MacKinnon C, Siegel R (eds) Directions in sexual harassment law. Yale University Press, New Haven, pp 651–677
Citron DK (2014) Hate crimes in cyberspace. Harvard University Press, Cambridge
Citron DK, Franks MA (2014) Criminalizing revenge porn. Wake Forest Law Rev 49:345–391
Cobb EP (2020) International workplace sexual harassment laws and developments for multinational for the employer. Routledge, Oxon
Conte A (2010) Sexual harassment in the workplace: law and practice, 4th edn. Wolters Kluwer, New York
Cusack S (2013) The CEDAW as a legal framework for transnational discourses on gender stereotyping. In: Hellum A, Aasen HS (eds) Women's human rights: CEDAW in international, regional and national law. Cambridge University Press, Cambridge, pp 124–157
Daniels J (2009) Cyber racism: white supremacy online and the new attack on civil rights. Rowman & Littlefield, Lanham
Daum CW (2009) Feminism and pornography in the twenty-first century: the internet's impact on the feminist pornography debate. Women's Rights Law Rep 30:543–565
Davidson A (2009) The law of electronic commerce. Cambridge University Press, Melbourne
Decew JW (1984) Violent pornography: censorship, morality and social alternatives. J Appl Philos 1:79–94
Dekeseredy W, Carlson AD (2020) Understanding the harms of pornography: the contributions of social scientific knowledge. Cult Reframed 2020:1–12
Delgado R, Stefancic J (2018) Understanding words that wound. Routledge, New York
Dripps D (1992) Beyond rape: an essay on the difference between the presence of force and the absence of consent. Columbia Law Rev 92:1780–1809
Dworkin A (1989) Pornography: men possessing women, 4th edn. Plume, New York
Dworkin R (1996) Freedoms law: the moral reading of the American constitution. Oxford University Press, Oxford
Dworkin R (1997) Liberty and pornography. In: Gruen L, Panichas GE (eds) Sex, morality, and the law. Routledge, New York, pp 223–231
Elbahtimy M (2014) The right to be free from the harm of hate speech in international human rights law. University of Cambridge, CGHR, Working Paper 7, pp 1–20. https://www.repository.cam.ac.uk/bitstream/handle/1810/245215/CGHR_WP_7_2014_Elbahtimy.pdf?sequence=1&isAllowed=y. Accessed 8 Feb 2022
Eriksson M (2011) Defining rape: emerging obligations for states under international law? Martinus Nijhoff, Leiden
Evans M (2006) What's morality got to do with it? The gender-based harms of pornography. S Cross Univ Law Rev 10:89–137
Evans M (2007) Censorship and morality in cyberspace: regulating the gender-based harms of pornography online. S Cross Univ Law Rev 11:1–58
Farber DA (1980) Civilizing public discourse: an essay on professor Bickel, Justice Harlan, and the enduring significance of Cohen v. California. Duke Law J 1980:283–303
Feinberg J (1985) Offense to others. Oxford University Press, Oxford
Fentonmiller KR (1994) Verbal sexual harassment as equality-depriving conduct. Univ Mich J Law Reform 27:565–611
Ferguson C et al (2022) Pornography and sexual aggression: can meta-analysis find a link? Trauma Violence Abuse 23:278–287
Fiss O (1996) The irony of free speech. Harvard University Press, Cambridge
Franks MA (2012) Sexual harassment 2.0. Md Law Rev 71:655–704
Gagliardone I et al for UNESCO (2015) Countering online hate speech. https://unesdoc.unesco.org/ark:/48223/pf0000233231. Accessed 11 Feb 2022
Gilden A (2016) Punishing sexual fantasy. William Mary Law Rev 58:419–491
Glomb T et al (1997) Ambient sexual harassment: an integrated model of antecedents and consequences. Organ Behav Hum Decis Process 71:309–328

Goldberger D (1991) Sources of judicial reluctance to use psychic harm as a basis for suppressing racist, sexist and ethnically offensive speech. Brook Law Rev 56:1165–1212
Harris DP et al (2005) Sexual harassment: limiting the affirmative defense in the digital workplace. Univ Mich J Law Reform 39:73–98
Heller B (2007) Of legal rights and moral wrongs: a case study of internet defamation. Yale J Law Fem 19:279–286
Henckaerts J-M, Doswald-Beck L, for the ICRC (2005) Customary international humanitarian law: vol I: Rules IHL. Cambridge University Press, Cambridge
Hughes D (2002) The use of new communications and information technologies for sexual exploitation of women and children. Hastings Women's Law J 13:127–146
Hunter N, Law S (1988) Brief amici curiae of feminist anti-censorship task force. Univ Mich J Law Ref 21:69–136
Keller SE (1993) Viewing and doing: complicating pornography's meaning. Geo Law J 81:2195–2242
Klaassen M, Peter J (2015) Gender (in)equality in internet pornography: a content analysis of popular pornographic internet videos. J Sex Res 52:721–735
Koltay A, Wragg P (eds) (2020) Comparative privacy and defamation. Edward Elgar, Cheltenham
Lambiase J et al (2017) Women versus brands: sexist advertising and gender stereotypes motivate transgenerational feminist critique. In: Golombisky K, Kreshel P (eds) Feminists, feminism, and advertising: some restrictions apply. Rowman & Littlefield, Lanham, pp 29–60
Langton R (1993) Speech acts and unspeakable acts. Philos Public Aff 22:293–330
Langton R (2011) Symposium on Langton's sexual solipsism. Jurisprudence 2:379–440
Leary MG (2008) Self-produced child pornography: the appropriate societal response to juvenile Self-sexual exploitation. Va J Soc Policy Law 15:1–50
Lederer L (ed) (1980a) Take back the night: women on pornography. William Morrow and Company, New York
Lederer L (1980b) Then and now: an interview with a former pornography model. In: Lederer L (ed) Take back the night: women on pornography. William Morrow and Company, New York, pp 57–70
Levin A (2010) The cost of free speech: pornography, hate speech, and their challenge to liberalism. Palgrave Macmillan, New York
Levmore S, Nussbaum MC (2010) Introduction. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 1–14
Lisby G (2004) No place in the law: the ignominy of criminal libel in American jurisprudence. Comm Law Policy 9:433–488
Livingstone S, Bober M for London School of Economics (2005) UK children go online: final report of key project findings. http://eprints.lse.ac.uk/399/1/UKCGO_Final_report.pdf. Accessed 21 Mar 2022
Long JH, Chen G-M (2007) The impact of internet usage on adolescent self-identity development. China Media Res 3:99–109
MacKinnon C (1979) Sexual harassment of working women: a case of sex discrimination. Yale University Press, New Haven
MacKinnon C (1984) Not a moral issue. Yale Law Policy 2:321–345
MacKinnon C (1985) Pornography, civil rights and speech. Harv Civ Rights Civ Liberties Law Rev 20:1–70
MacKinnon C (1993) Only words. Harvard University Press, Cambridge
MacKinnon C (1995) Vindication and resistance: a response to the Carnegie Mellon study of pornography in cyberspace. Geo Law J 83:1959–1968
MacKinnon C, Dworkin A (1985) Appendix: the MacKinnon/Dworkin pornography ordinance. William Mitchell Law Rev 11:110–125
MacKinnon C, Dworkin A (eds) (1997) In Harm's way: the pornography civil rights hearings. Harvard University Press, Cambridge
Mahoney KE (2010) Speech, equality, and citizenship in Canada. Comm Law World Rev 39:69–99

Maris C (2013) Pornography is going on-line: the harm principle in Dutch law. Law Democr Dev 17:1–23
Marshall E, Miller H (2019) Consistently inconsistent: a systematic review of the measurement of pornography use. Aggress Violent Behav 48:169–179
Martellozzo E et al (2020) Researching the affects that online pornography has on U.K. adolescents aged 11 to 16. SAGE Open, pp 1–11
Matsuda M (1989) Public response to racist speech: considering the victim's story. Mich Law Rev 87:2320–2381
McCartan K, McAllister R (2012) Mobile phone technology and sexual abuse. Inf Commun Technol Law 21:257–268
McGlynn C (2022) Cyberflashing: consent, reform and the criminal law. J Crim Law 2022:1–17
McGlynn C, Rackley E (2009) Criminalising extreme pornography: a lost opportunity. Crim Law Rev 4:245–260
McGlynn C, Rackley E (2017) Image-based sexual abuse. Oxf J Leg Stud 37:534–561
McGraw D (1995) Sexual harassment in cyberspace: the problem of unwelcome E-mail. Rutgers Comput Technol Law J 21:491–518
Mena P et al (2020) Misinformation on Instagram: the impact of trusted endorsements on message credibility. Social Media + Society, April–June 2020, pp 1–9
Mikkola M (2019) Pornography: a philosophical introduction. Oxford University Press, Oxford
Mill JS (1859) On liberty, 2nd edn. John W. Parker and Son, London
Mitchell KJ et al (2007) Trends in youth reports of sexual solicitations, harassment and unwanted exposure to pornography on the internet. J Adolesc Health 40:116–126
Murray A (2016) Information technology law: the law and society, 3rd edn. Oxford University Press, Oxford
Nair A (2007) Internet content regulation: is a global community standard a fallacy of the only way out. Int Rev Law Comput Technol 21:15–25
Nussbaum M (2010) Objectification and internet misogyny. In: Levmore S, Nussbaum M (eds) The offensive internet: speech, privacy, and reputation. Harvard University Press, Cambridge, pp 68–90
Oster J (2017) Which limits on freedom of expression are legitimate? Divergence of free speech values in Europe and the United States. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge, pp 39–47
Patton RB (2015) Taking the sting out of revenge porn: using criminal statutes to safeguard sexual autonomy in the digital age. Geo J Gend Law 16:407–444
Posner R (2006) Not a suicide pact: the constitution in a time of national emergency. Oxford University Press, Oxford
Powell A, Henry N (2016) Sexual violence in the digital age: the scope and limits of criminal law. Soc Leg Stud 25:397–418
Powell A, Henry N (2017) Sexual violence in a digital age. Palgrave Macmillan, London
Powell A et al (2020) Image-based sexual abuse: an international study of victims and perpetrators. Royal Melb Inst Technol 2020:1–15
Quenivet N (2005) Sexual offenses in armed conflicts and international law. Brill, Leiden
Reed C (2012) Making laws for cyberspace. Oxford University Press, Oxford
Rehof LA (1993) Guide to the Travaux Préparatoires of the United Nations Convention on the elimination of all forms of discrimination against women. Martinus Nijhoff, Dordrecht
Richardson-Self L (2018) Woman-hating: on misogyny, sexism, and hate speech. Hypatia 33:256–272
Rigaux F (2004) Abridged or forbidden speech: how can speech be regulated through speech? In: van Hoecke M (ed) Epistemology and methodology of comparative law. Hart, Portland, pp 285–298
Ringrose J et al (2012) A qualitative study of children, young people and 'sexting': a report prepared for the NSPCC. https://www.researchgate.net/publication/265741962_A_

qualitative_study_of_children_young_people_and_'sexting'_a_report_prepared_for_the_NSPCC. Accessed 21 Mar 2022
Ross JM et al (2019) Sexting coercion as a component of intimate partner polyvictimization. J Interpers Violence 34:2269–2291
Roth LM (1999) The right to privacy is political: power, the boundary between public and private, and sexual harassment. Law Soc Inq 24:45–71
Rowbottom J (2012) To rant, vent and converse: protecting low level digital speech. Camb Law J 71:355–383
Russell D (1980) Pornography and violence: what does the new research say? In: Lederer L (ed) Take back the night: women on pornography. William Morrow and Company, New York, pp 218–238
Sandmann W (1994) Three ifs and a maybe: Mari Matsuda's approach to restricting hate speech. Commun Stud 45:241–262
Schneider EM (1986) Describing and changing: women's self-defense work and the problem of expert testimony on battering. Women's Rights Law Rep 9:195–222
Schwartzman L (1999) Liberal rights theory and social inequality: a feminist critique. Hypatia 14: 26–47
Searles J (1994) Sexually explicit speech and feminism. Rev Jur U P R 63:471–492
Sekowska-Kozlowska K et al (2022) Sexist hate speech and the international human rights law: towards legal recognition of the phenomenon by the United Nations and the Council of Europe. Int J Semiot Law, pp 1–23
Shor E, Seida K (2019) "Harder and harder"? Is mainstream pornography becoming increasingly violent and do viewers prefer violent content? J Sex Res 56:16–28
Shor E, Seida K (2020) Aggression in pornography: myths and realities. Routledge, Oxon
Siegel R (2004) Introduction: a short history of sexual harassment. In: MacKinnon C, Siegel R (eds) Directions in sexual harassment law. Yale University Press, New Haven, pp 1–42
Sivakumaran S (2007) Sexual violence against men in armed conflict. Eur J Int Law 18:253–276
Sjöholm M (2017) Gender-sensitive norm interpretation by regional human rights law systems. Brill, Leiden
Slane A (2005) Home is where the internet connection is: law, spam and the protection of personal space. Univ Ottawa Law Technol J 2:255–290
Smith G (2007) Internet law and regulation, 4th edn. Sweet & Maxwell, London
Sparks B (2021) A snapshot of image-based sexual abuse (IBSA): narrating a way forward. Sex Res Soc Policy 2021:1–16
Strossen N (1996) Hate speech and pornography: do we have to choose between freedom of speech and equality? Case Western Res Law Rev 46:449–478
Sullivan K (1992) The first amendment wars. New Republic 207:35–40
Suzor N et al (2017) Non-consensual porn and the responsibilities of online intermediaries. Melb Univ Law Rev 40:1057–1097
Taylor JK (1994) Does sexual speech harm women – the split within feminism. Stanf Law Policy Rev 5:49–64
Temkin J (2002) Rape and the legal process. Oxford University Press, Oxford
Tsesis A (2002) Prohibiting incitement on the internet. Va J Law Technol 7:1–41
Tulkens F (2012) When to say is to do: freedom of expression and hate speech in the case-law of the European Court of Human Rights. In: Casadevall J et al (eds) Freedom of expression: essays in honour of Nicolas Bratza. Wolf Legal Publishers, Oisterwijk, pp 279–295
Vera-Gray F et al (2021) Sexual violence as a sexual script in mainstream online pornography. Br J Criminol 61:1243–1260
Volokh E (1996) Freedom of speech in cyberspace from the listener's perspective: private speech restrictions, libel, state action, harassment, and sex. Univ Chic Leg Forum 1996:377–436
Waldron J (2012) The harm in hate speech. Harvard University Press, Cambridge
West R (1993) Narrative, authority, and law. University of Michigan Press, Ann Arbor

West R (2000) The difference in women's hedonic lives: a phenomenological critique of feminist legal theory. Wis Women's Law J 15:149–215
Westen P (2016) The logic of consent: the diversity and deceptiveness of consent as a defense to criminal conduct. Routledge, New York
White AE (2006) Virtually obscene: the case for an uncensored internet. McFarland Co. Inc., London
Williams ML et al (2020) Hate in the machine: anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. Br J Criminol 60:93–117
Wright PJ, Funk M (2014) Pornography consumption and opposition to affirmative action for women: a prospective study. Psychol Women Q 38:208–221
Zavrsnik A (2010) Towards an overregulated cyberspace: criminal law perspective. Masaryk Univ J Law Technol 4:173–190

Chapter 5
Summary and Conclusion

5.1 Gendered Spheres

Internet regulation is contentious. The limited harmonisation in this field in part stems from an ideological discord, with varying approaches at the domestic and regional levels on the values and rights to be enforced, and who may or should intervene to restrain content and conduct. This in turn is informed by practical challenges related to control, in view of Internet architecture. Although not flawless—with abstract norms not specifically tailored to this environment—the regime of international human rights law is a valuable means of harmonising regulation offering protection to individuals against online transgressions, particularly considering the transborder character and effect of the Internet. This body of law provides a universal language of values; defines what is harmful; delineates the scope of rights and freedoms; concretises state obligations—including regulation of Internet intermediaries—and provides mechanisms for balancing in conflicts of rights, aspects which need development in the regulation of the Internet. Whereas states in the libertarian approach to cyberspace thus often are construed as passive observers—in contrast to users and intermediaries—it must be recognised that they are important players for the '. . .creation, maintenance, reinvention and reshaping' of the Internet.[1] While other actors, such as Internet intermediaries, share the governance of the online sphere, states maintain their international human rights law obligations to the same extent online as offline. Although increasingly categorised as a significant public sphere, the Internet accordingly does not acquire a special legal status in this regard.

In order to consider the effectiveness of international human rights law in the regulation of online gender-based harm, potential gaps, inconsistencies and disadvantages in this area of law have been considered in the book, in addition to the online/offline coherence in the application of provisions. In this regard, the feminist

[1] Warf (2017), p. 164.

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7_5

legal method of "asking the woman question" has been employed, which involves assessing whether law—in this case, international human rights law—is inclusive of women's distinct life experiences, values and biological differences. This presumes that women and men differ in certain regards, for example, evinced through empirical studies and feminist legal theories.

The Internet is an important forum for the fulfilment of a range of human rights, valuable to self-discovery, to resist and respond to gender-based offences and to participate in the political discourse for the advancement of gender equality. However, empirical studies indicate that women are disproportionately victimised by certain online offences, such as sexual harassment and image-based sexual abuse, and affected in different ways than men, for example, in relation to defamation and the non-consensual disclosure of personal information. In consideration of the disproportionate impact on women and the gendered nature of the harm—often involving sexualisation and objectification of women—such offences constitute gender-based violations. That is, statistics clearly indicate that women's experiences on the Internet, in certain respects, is distinct. From a feminist theoretical perspective, difference may stem from biological distinctions between men and women, structural power hierarchies or failures to ensure formal equality, depending on the theoretical approach. In view of the nature of the offences analysed in the book, difference has mainly been approached from a constructivist perspective.

In considering whether law recognises such differences, the construction, aim and effect of social spheres, areas of law and legal principles have been assessed in relation to women, with a focus on gender. In this book, it involved the architecture of the Internet, the structure of public international law, the harm principle as employed in domestic law and international human rights law as well as the scope of particular rights and obligations, including regulation of liability. This indicated that while the design and ideology embodied by the Internet undermine constraints of harmful behaviour and exacerbate certain forms of gender-based offences, structural challenges in effectively regulating gender-based offences online arise also through the foundation and boundaries of international human rights law.

The primarily male development, content and use of the Internet entails that, in several regards, it can be categorised as a gendered space, adapted to suit male objectives. Gender roles and offences are, as a result, shaped by and, in turn, inform the medium of communication, such as the design of technology. Whereas the root causes of gender-based offences include gender stereotyping and the social, legal and political disempowerment of women, technology affects the form and prevalence of crime, for example, evident through the development of image-based sexual abuse and doxing. Criminal behaviour may also benefit from technology. For example, stalkers can locate individuals through social media geolocations, perpetrators may remain anonymous, and the automation of services ensures easy and low-cost dissemination, for example, of obscene pornography or live sexual abuse.

Whereas the social dynamics and gender hierarchies that exist IRL are replicated online, they are also exacerbated through the architecture of the Internet, underpinned by ideologies of libertarianism. The regular constraints on human behaviour, including law and social norms, are undermined and transformed online.

Its decentralised nature—through its end-to-end design—is central in this regard. The governance of this sphere by Internet intermediaries and online media publishers limits state control over content and communication. Jurisdictional complexities and a lack of regulation at both the domestic and international level further undermine governmental control. Meanwhile, user anonymity impedes investigation and prosecution of offenders. These technical features—particularly anonymity—enhance social norms detrimental to gender equality and facilitate certain types of harm by encouraging coarser language, mob-like behavior and radicalised views. In addition, the universal availability, durability and rapid means of dissemination and asynchronicity of information may aggravate harm. In this regard, it is clear that international regulation of the Internet cannot consider the environment solely in terms of its technical features, but also how technology interacts with and affects gender relations, for example, through the exacerbation of harmful gender stereotypes.

International human rights law has also been categorised by feminist legal scholars as a male space in consideration of its theoretical foundation, structure and content, which in certain instances impede the development of regulation corresponding to women's life experiences and values. This includes the *de facto* delineation of public/private spheres of regulation, which excludes the direct liability of non-state actors and has generated a male-oriented list, hierarchy and content of rights. This has resulted in codification gaps of gender-based harm against women in general and limited attention given to *online* gender-based offences, with a focus on terrorism, data protection and child abuse as cyber offences in international law. Although the process of developing and interpreting international human rights law provisions is increasingly gender-sensitive and the public/private distinction to a degree has been pierced through the construction of positive state obligations to prevent interpersonal harm, issues still remain in terms of subjects of liability and the gendered content of rights. While certain attributes of international human rights law place obstacles to the acknowledgment of women's rights *per se*, others particularly affect regulation of online offences. Pre-existing issues in international human rights law—concerning the gendered regulation, content and hierarchy of rights—are also exacerbated by the Internet, with gender-based online harm often falling outside the scope of regulation or overridden in conflicts with the freedom of expression.

Whereas an online/offline coherence has been affirmed at an abstract level in international human rights law—applicable also to gender-based offences—the online environment in practice affects legal assessments in several regards. Coherence does not necessarily entail that no adaptation is made in the application of rights, but that the aims, values and effectiveness of rights are maintained. Such a contextual approach to the interpretation of rights is not only promoted by feminist legal theories—requiring consideration of gender and space—but also by the objectives of international human rights law, that is, ensuring effective protection of human rights. Through interpretation of abstract principles—such as human rights law provisions—facts, circumstances and contexts are consequently provided prominent weight. If not, the balancing of rights and interests becomes predisposed to certain results. A contextual approach thus evaluates the relationship between the

law, the Internet and gender: either maintaining the *status quo* of the scope of rights, adapting interpretations and/or requiring development of Internet architecture to align with human rights law norms. In doing so, the emphasis rests on ensuring substantive equality on the Internet, which increases the demands for a more responsive state.

Although existing treaty provisions have been applied by international and regional human rights law bodies in a manner that incorporates expressions and acts on the Internet, this has been done in an *ad hoc* manner. How rights should be transposed to the online context, in terms of their content, scope and the balancing in conflicts or restrictions of rights, has thus not been cohesively formulated. Nor has this been addressed in relation to the values of the law. For example, it is necessary to consider how the Internet undermines or heightens the protection of values, such as the democratic function of the freedom of expression or secrecy and sexual autonomy as aspects of the right to privacy. Beyond such broad issues, it is also clear that the Internet affects the delineation of concepts, be it actors—such as journalists, the media, publishers and public/private figures—or offences—including sexual violence and harassment. Mainly, contextual elements have been considered in the evaluation of harm and the scope of positive obligations for states in regulating liability. The latter has been addressed through proportionality assessments of state restrictions of rights and in the balancing in conflicts of interests, involving the freedom of expression *vis-à-vis* the right to privacy.

5.2 The Scope of Rights: Values, Harm and Balancing

Online gender-based offences are both generic and contextually specific, a standpoint affirmed by a range of international and regional human rights law bodies. Although the Internet has generated new types of offences, such as image-based sexual abuse, in view of their nature and the values transgressed, they may thus in many instances be addressed within existing rights and concepts. This includes both specific regulation on gender-based violence and gender stereotyping as well as general provisions encompassing such forms of harm, such as the right to privacy. Online harm against women is part of the broader spectrum of gender-based violations and the means of its elimination are consequently similar. Nevertheless, although no in-depth consideration of how the Internet affects international human rights law *vis-à-vis* such offences has been made, it is clear from the analysis that theoretical and practical challenges arise in the application of such provisions.

The explicit regulation or gender-sensitive interpretation of rights relevant to online gendered offences is dependent on the recognition of the severity of harm of conduct and content. The ideological foundation of the concept of harm thus affects the acknowledgment of online gender-based offences as violations of domestic law and international human rights law. In view of the nature of online offences, this primarily concerns the approach to the harm of speech—including its group-based effects—in addition to the Internet as the site for abuse.

The scope of the freedom of expression and legitimate restrictions delineate which forms of speech are harmful and informs the approach to the degree and nature of harm. Clarification of these aspects can be achieved not solely through considering the current interpretation of rights in international human rights law, but also rights theories. The latter allows for a critical examination of the potentially gendered nature of the underlying values of rights and how such values are integrated into Internet architecture or regulation. A predisposition towards certain values in law and architecture was in this regard noticed, undermining the regulation of online gender-based offences. Whether aiming to ensure individual autonomy or democracy, legal theories and international human rights law alike offer extensive protection of the freedom of speech, with restrictions as rare exceptions. In many respects, speech is in international human rights law considered either not harmful or—if causing hurt—as less harmful than restrictions on the freedom of expression. Accordingly, the regulation of gender-based harm requiring restrictions of the freedom of expression, such as sexual harassment, sexist hate speech and harmful pornography, is not particularly well-developed in international human rights law, even beyond the context of the Internet.

The approach to the harm of speech in turn affects its categorisation as a violation of a particular right in international human rights law and its place in the (informal) hierarchy of rights. At a general level, a hierarchy related to the physical or psychological nature of the harm is in place. The latter form of harm is mainly categorised as invoking provisions lower in the hierarchy, for example, qualified rights such as the right to privacy, in contrast to the prohibition on torture. This also affects the content of obligations. For instance, states must adopt domestic criminal laws prohibiting sexual violence—mainly interpreted to involve physical injury—while offences such as defamation and the non-consensual disclosure of personal information primarily require civil law remedies. Meanwhile, empirical studies indicate that victims of online offences mainly suffer psychological harm, as these are generally speech-based. With a hierarchy of physical harm considered more severe than psychological, both at the domestic and international level, online offences are thus *per se* categorised as less severe. The categorisation of speech as less harmful in turn weakens the development of transnational agreements on restricting online behavior.

Beyond the physical/speech-based dichotomy, the traditional approach to harm involves a focus on individual harm, as opposed to group-based injury. Although support for the acknowledgment of social harm can be garnered both from empirical studies and theoretical discourses, the former is limited. Rather, feminist legal theories—such as dominance feminism—as well as linguistic theories, affirm a correlation between speech, acts and gender inequality. Speech is in such frameworks considered a means of exercising power, constructing and maintaining social gender hierarchies.

Whereas international human rights law to a certain extent recognises group-based harm, for example, in relation to hate speech, it appears that the lack of a clear causality between speech and inequality hampers the categorisation of certain forms of harm as violations. This is also linked to the theoretical dispute on individual

agency—for example, *vis-à-vis* harmful pornography—where liberal theory influences the autonomy ascribed to individuals. Nevertheless, the correlation between speech and inequality is increasingly affirmed in international human rights law, *inter alia* by the United Nations Committee on the Elimination of Discrimination against Women (CEDAW Committee). Furthermore, the link between gender stereotypes and gender-based violence is well accepted in international human rights law. Where material exacerbates gender stereotypes, a connection to gender inequality can thus more easily be established. For instance, the CEDAW Committee (in concluding observations on pornography) and the CoE and the EU (in relation to sexist hate speech) recognise a link between stereotyped roles of women in the media and gender discrimination. Nevertheless, the causality is intermittently speculative and presumed in these sources, which could be strengthened through a more pronounced theoretical approach. This is particularly pertinent in the context of the Internet, in view of the widespread distribution and easy access to online material that exacerbate gender stereotypes or is otherwise harmful to women as a group.

The context of the Internet also has an impact on assessments of the nature and gravity of harm. The architecture of the Internet embodies the values of democracy and individual autonomy of the freedom of expression. This includes its decentralised structure—such as its end-to-end communication—and possibilities for individuals to affect its design and create and restrict content. User anonymity also limits possibilities of regulating speech, *de facto* aligned with the protection of individual autonomy. Although valuable in terms of audience rights, the Internet is thus particularly important from a speaker's perspective, as individuals are monitored in many other fora, be it in print press or broadcasting. As social norms on the Internet are aligned with a generous freedom of expression, harmful and offensive speech is thus accepted to a higher degree than IRL. This has affected the interpretation of the scope of the freedom of expression in this context. In fact, the ECtHR has in several cases involving defamation indicated that online speech is less harmful, primarily in view of social norms in this sphere. The *prevalence* of disinformation and vulgar and offensive speech accordingly affects the assessment of the harmful impact of statements. The Court has further noted that vulgar language may serve a stylistic purpose and style is protected as an aspect of communication alongside the substance of ideas. Furthermore, the distance between the perpetrator and victim, the asynchronicity of information and the possibility for victims to reply have also been raised as mitigating factors at both the domestic and international level. Thus, the Internet as the site for abuse has in this regard lowered the protection against such offences.

Beyond this general approach to the harm of online speech, the assessment of whether speech constitutes a particular offence, including hate speech, harassment or defamation, considers such factors as the identity of the speaker, the audience, content and form. In this regard, Internet design challenges the evaluation of the intent and effect of speech, given the lack of tonal and facial cues as well as a, frequently, global, heterogeneous audience. The tenor of posts may thus be evinced from the website in question, previous discussions or through such features as

emoticons and multimodal expressions. The specific pocket of the Internet thus also needs to be considered in the assessment of harm.

Meanwhile, the right to privacy in the main aims to protect individual autonomy—also in the social sense—and includes protection of sexual autonomy, dignity and reputation. The design of the Internet and social norms prevalent online to an extent enrich values such as sexual autonomy and identity explorations, aided by user anonymity. Simultaneously, these protected interests are frequently transgressed through gender-based online offences. Noted *inter alia* by the ECtHR, the risk of harm to reputation is elevated on the Internet, given the ease in disseminating defamatory statements and private images or other personal information without consent. Meanwhile, online sextortion and image-based sexual abuse involve intrusions on sexual autonomy.

Again, both the ideological foundation of the right to privacy and technological design affect the transposition of the right to the Internet. The scope of privacy *per se* is in fact contextual since the right is reflective of social norms. This includes assessing the scope of privacy in a sphere that can be categorised as public with private pockets. Although individuals maintain a right to privacy also in public areas, it is not as extensive and is relative to the person's reasonable expectation of privacy. The fluidity concerning the public/private nature of the Internet is evident in the approach by the ECtHR, which in cases on defamation and disclosure of private information has considered the particular website, privacy settings and number of visitors to determine the scope of the audience and thus the level of harm experienced. Publicity accordingly heightens harm to reputation and may even be a required aspect of a particular offence, such as in relation to the non-consensual disclosure of personal information.

Nevertheless, what level of privacy is reasonable in the context of the Internet has only to a limited extent been addressed in international human rights law. In consideration of Internet design and online social norms, theoretical frameworks in other disciplines—such as sociology and psychology—are useful in developing subjective and objective criteria. The private lives of individuals are increasingly displayed online, both consensually and non-consensually. While it may be argued that the willingness to share personal information reduces the scope of the protection of privacy, it is evident that the traditional focus on protecting secrecy cannot be fully applied online. This requires a clearer delineation of where and how privacy is to be protected online, with a focus on individual integrity rather than solely maintaining confidentiality. As noted, the ECtHR has in cases on the re-distribution of personal information held that, although protection is reduced once secrecy is lost, obligations to safeguard a person's dignity remain.

Furthermore, the right to privacy encompasses protection of physical integrity and sexual autonomy. In this regard, the application of existing legal concepts—mainly construed as encompassing physical acts—to the Internet presents challenges, given that violations primarily are conducted through speech. The distinction between speech and conduct is not self-evident, particularly on the Internet, where the online and physical worlds interact. For example, whether sexual harassment constitutes a form of sexual violence or merely offensive speech has an impact on

whether the conduct is considered within the discourse on the freedom of expression—potentially subject to the balancing of interests—or as a violation of sexual autonomy, with stricter obligations for the state. Furthermore, while a positive obligation to criminalise sexual violence is widely acknowledged—affirmed also in relation to the Internet—its applicability to speech-based interactions online is unclear, given the novelty of the issue. The lack of physical acts involved in online sexual offences confronts the perception and legal boundaries of sexual violence. This entails that the transposition of definitions of violations or specific elements of offences, such as the *actus reus* of rape, risks excluding online acts from the scope of protection. It is thus important that online transgressions that violate the sexual autonomy of individuals are categorised as sexual violence and that particular offences are defined in a manner that encompasses such harm. It is also imperative that non-consent in relation to sexual violence is interpreted in light of means of coercion through new technologies.

Moreover, Internet architecture and the great value ascribed to the Internet appears to influence the cost-benefit analysis in the balancing of rights and interests. In proportionality and balancing exercises, primarily by the ECtHR, a predisposed favoring of protecting the freedom of expression is notable in certain instances, particularly in relation to the Internet, which stems from the fundamental values attached to this right and the Internet as a public or quasi-public sphere. Both the freedom of expression and the Internet are regarded as not only important to democracy but also for the fulfilment of other human rights. In view of this, states maintain a wide margin of appreciation in terms of regulation. The value to the public good is thus perceived as greater than, at least, minor forms of individual harm. Where the functions of the Internet are threatened, individual interests are forfeited in order to protect the architecture and effective operation of the Internet. With ICT-based offences dismissed as harmless teasing, the balancing between the protection of individuals and an unrestrained Internet is more likely to tip in favour of the latter. This entails that the impetus at the domestic and international level to regulate online conduct is undermined. While safeguards against the overregulation of speech is essential—such as restrictions on blocking, censoring and monitoring—the values associated with the Internet, such as democracy, must also be approached as producing broader obligations for states to ensure substantive equality. This must also be considered in balancing exercises. Gender equality in this regard pertains to both the reduction of the digital gender gap—including the development of Internet infrastructure, enhancing physical access and strengthening women's skills and education—as well as the regulation of online content to ensure protection against gender-based violations and harmful gender stereotyping.

These theoretical approaches to the concept and assessment of harm, and the degree of state involvement in the private sphere, have concrete consequences for the regulation of online gender-based offences, creating or widening gaps and inconsistencies in terms of rights and obligations. Whereas these issues have long been identified as problematic by feminist legal scholars *vis-à-vis* women's human rights in general, it is clear that new challenges arise in relation to Internet regulation. The narrow approach in international human rights law to sexual harassment, the

gendered delineation of hate speech and the limited regulation of harmful pornography constrains the possibilities of addressing harmful online content. As these forms of harm are widespread and made mainstream online—reproducing and normalising gender stereotypes—the disinterest in this area of law has gendered effects in terms of protecting online users. A call for more explicit regulation or an extension of the scope of rights can thus been made in relation to such content. Meanwhile, the protection against defamation and the non-consensual disclosure of personal information is clearly a part of international human rights law, albeit to a more limited extent involving positive obligations. However, in relation to these violations, the context of the Internet has affected the assessment of harm in a manner disadvantageous to the recognition of online offences. Also the prohibition of sexual violence is unambiguous in international human rights law. Nevertheless, the current approach to the definition of concepts and the hierarchy of various sexual offences impede the recognition of online harm to sexual autonomy. A contextual approach thus cannot be reduced to solely considering the sphere of transgressions but also the implications *vis-à-vis* particular social groups, such as women. A re-thinking of the theoretical approach to harm and the values of speech—which in turn affects balancing exercises—is consequently essential in effectively addressing online offences in international human rights law.

5.3 Obligations and Liability

Internet architecture must also be borne in mind in terms of the content of state obligations, mainly affecting the regulation of liability. The Internet has an impact on the scope of state obligations at a general level, with several factors relevant to the regulation of online gender-based offences generating a broad margin of appreciation for states in the choice of means. For example, this position has been taken by the ECtHR in cases of conflicts of rights. As noted, the Internet fuels conflicts between, for example, the freedom of expression and the right to privacy, or intra-right conflicts in terms of values or beneficiaries. A broad leeway is also provided to states in the choice of measures to fulfil positive obligations. More specifically, issues affecting the operation of the Internet generate a wide scope, given its importance as a (quasi) public sphere. The latter aspect has also been affirmed in other international human rights law sources. Furthermore, restrictions of rights on the basis of public morals involve a wide state deference, such as in relation to pornography. Consequently, states are in many regards given a broad scope in choosing the means of regulating the Internet. Nevertheless, where restrictions involve intimate aspects of a person's life as well as particularly grave offences, the scope of the margin of appreciation is narrowed, for instance, with regard to sexual violence.

In relation to interpersonal offences, states mainly incur positive obligations to protect individuals. Prevention in the form of legislation—delineating individual responsibility—and effective investigation and prosecution of offences, is central.

This remains the main approach on the Internet. As concluded, the adoption of effective criminal laws is an *obligation* for states in order to protect individuals against sexual violence, extending beyond physical acts, which has been affirmed also in the online context. In relation to other mainly speech-based offences, the content of obligations is less clear. Whereas, for example, the ECtHR for the most part has addressed such from the viewpoint of negative obligations, that is, whether state restrictions on speech are legitimate, obligations to prohibit speech arise in certain instances. For example, obligations to prohibit hate speech and speech inciting violence can be found in multiple international sources, including the ICERD. Treaty bodies such as the CEDAW Committee and the UN HRC place obligations on states to regulate harmful pornographic material. In contrast, the ECtHR permits domestic restrictions of such material, for example, to protect morals or the rights and freedoms of others, the first aim engendering a broad margin of appreciation. Nevertheless, obligations to ensure access to justice for victims of defamation, the non-consensual disclosure of private information and harassment have been affirmed. However, although criminal law regulation has been considered appropriate in certain instances, speech-based offences primarily require civil law remedies. As state obligations are more extensive in the protection against sexual violence than, for example, defamation, it is consequently important whether offences such as the non-consensual receipt, recording or distribution of intimate images is categorised as sexual violence at the international level.

International human rights law generally does not direct states in the exact formulation of domestic law, with due regard of the margin of appreciation states. Limited guidance is thus provided in the formulation of domestic laws pertaining to online offences, be it the elements of crimes or the delineation of modes of individual liability. In the main, it is considered sufficient that states prohibit the offence in question. Nevertheless, on rare occasions, regional human rights law courts and UN treaty bodies indicate elements of crimes required to ensure the *effective* implementation of rights, for example, in relation to the definition of rape. Narrowly defined domestic criminal or civil laws excluding digital offences may thus contravene international human rights law by not ensuring individual protection. Although international human rights law does not provide clear guidance on the issue, it includes the lack of regulation of threats to distribute intimate images or other forms of image-based sexual abuse, sexist hate speech and online sexual harassment. The Internet as a medium also brings complexities in relation to individual liability. More specifically, the Internet provides new means of perpetrating crimes where, for example, coercion of physical acts occurs digitally. This requires consideration of appropriate forms of liability in relation to certain offences, such as sexual violence or sexual harassment. It is generally a matter for domestic criminal law but it is also an aspect in ensuring that criminal laws effectively prevent interpersonal violence, and thus an element of the positive obligations of states.

The Internet also affects obligations to conduct effective investigations. While user anonymity—affirmed as a human right—and restraints on surveillance and the collection of personal information impede effective investigations and limit possibilities of redress, the establishment of laws and procedures allowing for the

identification of perpetrators is, at a minimum, an obligation for states in relation to certain grave offences, such as sexual violence and hate speech, while adhering to procedural guarantees. As the disclosure of data is balanced against the rights of the anonymous user, as well as the freedom of expression of media publishers, this will be determined on a case-by-case basis, and may also encompass, for example, defamation. Accordingly, in the balancing in conflicts of rights, interests such as protecting anonymity may be considered supreme to obligations to investigate, affirmed in relation to political speech in the case law of the ECtHR.

Technical difficulties in identifying perpetrators, the extensive control of the Internet by intermediaries as well as the limited presence of state authorities mean that liability for online content must necessarily be distinct in certain regards. Internet intermediaries and media publishers, in different ways, provide the means for perpetrating online violations. Given the narrow possibilities of holding individual perpetrators accountable, placing secondary liability on such entities is thus increasingly viewed as a practical approach to Internet regulation. Nevertheless, this pertains to certain types of offences and particular pockets of the Internet. When involving content in private fora, or conduct merely initiated through the Internet but perpetrated offline, liability generally rests on the offender.

Intermediaries are currently subject to limited regulation at the international level. Mainly soft law norms and self-regulation guide their respect for human rights law. The development of an international treaty for businesses and human rights has been proposed but has yet to be developed. Meanwhile, the e-Commerce Directive in EU law establishes frameworks for liability exemptions, whereas the proposed Digital Services Act (DSA) will place obligations on intermediaries to remove illegal material upon notification. Currently, this does not encompass gender-based violations. However, the intended EU directive on violence against women aims to harmonise domestic laws on rape, cyber harassment and incitement to violence against women, although narrowly defined. While Very Large Online Platforms (VLOPs) will acquire certain obligations to address harmful content—including, for example, harmful gender stereotyping—these are limited to due diligence measures.

State obligations *vis-à-vis* such entities are also developing in international human rights law. As viewed, state obligations to regulate the secondary liability of primarily media publishers arise, given the level of control these corporations exert over online content. In the main, this involves hate speech and direct threats against individuals, in effect requiring pre-monitoring and immediate removal of content. In relation to intermediaries, such as blogs and search engines, limited forms of liability have also been accepted in relation to defamation, such as the use of notice-and-takedown mechanisms. Given the character of the case law of the ECtHR, this has been reviewed in terms of the legitimacy of restrictions, rather than as obligations, albeit the Court in *obiter dicta* has referred to obligations under international law. Additional obligations arise in relation to the right to be forgotten. Given that sexist speech has yet to be included affirmatively within the scope of "hate speech", beyond the proposed EU directive and soft law, it is thus clear that limited obligations are currently placed on intermediaries and online media

publishers in relation to gender-based online offences. This again indicates that gender-based harm is considered less severe.

In terms of tools for preventing and responding to harmful or illegal material, the use of blocking, filtering and censoring is restrained by international human rights law, which affects the possibilities of states to control content, such as violent pornography and hate speech. Beyond notice-and-takedown mechanisms in relation to certain forms of unlawful content, the ECtHR has in view of Internet architecture placed the responsibility mainly on users to, for example, filter certain types of content (e.g. spam). The use of age verification tools by website operators has also been suggested as a means of limiting the availability of content that is mainly deemed harmful to children, for example, pornography. Given the narrow approach to "illegal" (EU law) and "clearly unlawful" (ECtHR) speech, the main impetus is thus currently on users to restrict and flag offensive or inappropriate material.

Regulation of online content is in this regard guided by a practical approach, where values are balanced against the technical abilities to control content, and the financial and social impact of constraints. This includes the social benefits and disadvantages of precautionary measures, instruments available to prevent harm and actual costs for online platforms, relative to the type of illegal material. Beyond the distinction between media publishers and intermediaries, a graduated and differentiated approach is accordingly applied in international human rights law and EU law, considering the particular qualities of the platform in question. What types of content intermediaries are able to assess in terms of legality and whether and how such may be controlled is also considered. As noted, speech-based offences often require a contextual assessment. Complexity, for instance, arises in evaluating consent in relation to the publication of intimate images. Similarly, assessing the contextual nuance of the harm of speech potentially categorised as hate speech and defamation, as well as what types of speech fall within the scope of certain offences, such as sexual harassment, is challenging. Even threats which, to a degree, involve more objective standards, must be assessed in view of, for instance, the specific website, the preceding dialogue and the relationship between the author and the victim. These exercises not only require advanced legal evaluations aligned with domestic law and, potentially, international human rights law, but also an awareness of the circumstances in the particular case. Although this may lead to more objective standards in practice, contextual considerations are necessary components of these offences. Although monitoring by intermediaries primarily involves a risk of encroaching upon the freedom of expression through removing legitimate speech, as noted previously, it is simultaneously a gendered process, with online sexist speech and harassment less likely to be flagged or deleted. Not only may thus the social benefits of the freedom of expression outweigh the level of harm of gender-based offences in balancing exercises, but the perceived complexity in controlling such content is a factor. As noted, sexist speech is common in everyday language, which affects the cost-benefit analysis. However, such practical concerns cannot guide the approach to liability. Rather, enhanced and more nuanced state obligations to regulate intermediary/media publisher liability is warranted in relation to gender-based offences, which requires a close collaboration between such entities and states.

However, in order to develop effective domestic and international regulation on this topic, a broader approach to obligations is necessary. As noted in Sect. 2.3 on the gendering features of the Internet, this sphere undermines regular constraints on human behaviour, including law, architecture and social norms. In view of this, and given the structural causes of gender-based violations, state obligations must also address the underlying factors producing gendered digital harm. This may include non-legal measures such as public awareness campaigns and education on gender stereotypes, gender roles and the Internet. Particularly in the context of the Internet, it has been noted that social norms play an important role in the control of behaviour. In instances of a gap between established social norms and legal provisions, domestic laws have largely been ineffective. At the same time, international human rights law also has an expressive function, affecting social norms and behaviour. As noted, the non-regulation of the Internet promotes the libertarian approach of a sphere free from state intervention, which exacerbates gender-based harm in particular.

Similarly, the advancement of liability regimes can be viewed as a means of influencing the adaption of technological design. In certain instances, law appears to conform to technology, such as in relation to hate speech, with the ECtHR indicating a turn towards more objective assessments in relation to online media publisher liability. Meanwhile, increasingly advanced technology is being developed, for instance, to remove intimate images uploaded without the consent of the subject, to lexically detect hate speech and to ensure that users have greater control over the dissemination of and responses to published content. Although the alignment of ICTs with international human rights law is not generally construed as an obligation, there is potential for a gender-sensitive synchronisation of law and technology. Internet design may accordingly adapt to human rights law standards, with an aim of preventing gender-based violence and gender stereotypes. Meanwhile, the prevalence of harmful material online may also compel international human rights law to develop clearer regulation, for example, on harmful pornography. The construction of broader and more gender-sensitive liability regimes would thus not only bring greater redress for victims, but also encourage the technological development of more effective means of detecting and removing harmful content.

In conclusion, from the viewpoint of feminist theories—be it cyberfeminism or the feminist critique of international law—the architecture of the Internet and the content of international human rights law are not inevitable or fixed but reflect the ideologies of their creators. While this non-neutrality has produced a legal regime ill-equipped in addressing online harm against women, the cyberfeminist approach that technology is shaped by social interests and gender relations entails that technology can also be redesigned to achieve certain public goals. Meanwhile, there is potential for the development of gender-sensitive and contextually specific international obligations for states in this area—including regulation of intermediary liability—aiding the advancement of a gender equal space, through decreasing impunity and encouraging the alignment of Internet architecture with international human rights law.

Reference

Warf B (2017) Alternative geographies of cyberspace. In: Kohl U (ed) The net and the nation state. Cambridge University Press, Cambridge

Index

A
ACHPR, *see* African Charter on Human and Peoples' Rights (ACHPR)
ACHR, *see* American Convention on Human Rights (ACHR)
African Charter on Human and Peoples' Rights (ACHPR), 36, 49, 98, 205, 212, 222, 238, 240, 241
African Commission on Human and Peoples' Rights, 31, 36, 39, 107, 140, 206, 208, 212, 213
American Convention on Human Rights (ACHR), 7, 33, 106, 112, 125, 257
Anonymity
  freedom of expression, 145
  international human rights law, 144
  internet, 61
Autonomy
  freedom of expression, 102
  gender equality, 21
  pornography, 315
  privacy, 123
  sexual violence, 205

B
Beijing Declaration, 2, 29, 31, 43, 97, 238, 241, 332
Belém do Pará Convention, 36–38, 41, 45, 49, 98, 205–207, 238, 240, 241
Blocking, 18, 27, 60, 78, 80, 113, 157, 176, 181, 182, 186, 280, 350, 354
Budapest Convention, 11, 19, 32, 110, 147, 158, 162, 177, 182, 287, 288, 290, 291, 293, 295, 299, 307, 325, 333
Bullying, 48, 63, 78, 134, 164, 243

C
CAT, *see* UN Committee Against Torture
CEDAW, *see* UN Convention on the Elimination of All Forms of Discrimination against Women; UN Committee on the Elimination of Discrimination against Women
CERD, *see* UN Committee on the Elimination of Racial Discrimination
CESCR, *see* UN Committee on Economic, Social and Cultural Rights
Children
  child-abuse, 32
  privacy, 134
  sexual violence, 208
  vulnerability, 11
CJEU, *see* Court of Justice of the European Union (CJEU)
CoE, *see* Council of Europe (CoE)
CoE Convention on Cybercrime, *see* Budapest Convention
CoE Convention on Preventing and Combating Violence against Women and Domestic Violence, *see* Istanbul Convention
Conflicts of rights, 138–142
Contextual approach
  feminist theories, 10
  international human rights law, 34

M. Sjöholm, *International Human Rights Law and Protection Against Gender-Based Harm on the Internet*, https://doi.org/10.1007/978-3-031-15866-7

Council of Europe (CoE), 1, 7, 9, 11, 12, 19, 27, 32, 36, 42, 47, 58, 98, 111, 151, 158, 164, 205, 209, 219, 225, 226, 265, 287, 306, 309, 322, 325, 329
Court of Justice of the European Union (CJEU), 177, 178
CRC, *see* UN Convention on the Rights of the Child
CRPD, *see* UN Convention on the Rights of Persons with Disabilities
Cultural feminist theory, 22
Cyberfeminism, 8, 55, 117, 355
Cyberlibertarianism, 17, 19, 80
Cyber norms, 60, 61, 65, 66, 69, 70, 75, 102, 114, 128

D
Defamation
  equality, 256
  freedom of expression, 258
  internet, 263–269
  privacy, 257
Deindividuation, 62
Democracy, 2, 4, 10, 17, 18, 20, 27, 31, 70, 79, 80, 84, 86, 96, 99, 102–107, 111, 112, 114, 117–121, 124, 130, 140, 141, 145, 146, 153, 161, 185, 245, 256, 269, 283, 286, 289, 294, 296, 297, 313, 347, 348, 350
Disclosure
  images, 210
  internet, 131–135
  right to privacy, 126, 271–284
Doxing, 48, 50, 134, 272, 275, 283, 344

E
ECHR, *see* European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)
ECJ, *see* European Court of Justice (ECJ)
e-Commerce Directive, 155, 160, 161, 163, 169, 177, 190, 191, 193, 273, 353
ECtHR, *see* European Court of Human Rights (ECtHR)
Equality
  feminist theories, 21
  gender-based violations, 38
  international human rights law, 23
  internet, 25
EU, *see* European Union (EU)
European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), 7, 44, 83, 94, 109–112, 127, 139, 147, 168, 215, 257, 273, 291, 294, 295, 299, 310, 313, 316
European Court of Human Rights (ECtHR)
  access to Internet, 26
  balancing, 140
  defamation, 257
  disclosure, 273
  equality, 24
  freedom of expression, 105
  harassment, 246
  hate speech, 289
  intermediary liability, 167–175
  pornography, 325
  right to privacy, 123
  sexual violence, 211
European Court of Justice (ECJ), 161, 162, 177, 178, 183, 184, 191
European Union (EU), 3, 49–52, 67, 77, 125, 157, 165, 184, 235, 273
Evolutive treaty interpretation method, 33, 213, 291, 310

F
Feminist legal methods, 7
Forced masturbation, 217–224
Forced nudity, 217–224
Freedom of expression
  international human rights law, 105
  internet, 112
  rights theories, 102

H
Harassment
  defamation, 255
  definition, 234
  online, 243
  sexual, 236–250
  threats, 250
Harm
  concept, 82, 83
  equality, 86–92
  feminist theories, 87
  international human rights law, 92–99
  linguistic philosophy, 88
  speech, 84–92
  technosocial, 77–82

Hate speech
equality, 285
freedom of expression, 286
harm, 284
international human rights law, 288–299
internet, 300–302
sexist, 303–314
Hyperlinks, 264, 270

I
IACmHR, *see* Inter-American Commission on Human Rights (IACmHR)
IACtHR, *see* Inter-American Court of Human Rights (IACtHR)
ICC, *see* International Criminal Court (ICC)
ICCPR, *see* International Covenant on Civil and Political Rights (ICCPR)
ICERD, *see* International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)
ICESCR, *see* International Covenant on Economic, Social and Cultural Rights (ICESCR)
ICL, *see* International criminal law (ICL)
ICT, *see* Information and Communication Technologies (ICTs)
ICTR, *see* International Criminal Tribunal for Rwanda (ICTR)
ICTY, *see* International Criminal Tribunal for the Former Yugoslavia (ICTY)
ILO, *see* International Labour Organization (ILO)
Image-based sexual abuse, 11, 46, 48, 53, 67, 77, 89, 125, 136, 165, 194, 204, 205, 209–211, 214, 216, 217, 226, 229, 231, 234, 241, 253, 272, 284, 344, 346, 349, 352
Information and Communication Technologies (ICTs), 2, 20, 113, 211, 303
Inter-American Commission on Human Rights (IACmHR), 2, 26, 28, 29, 31, 38, 98, 106, 109, 111, 113, 118, 121, 123, 140, 145, 156, 160, 181, 213, 219, 226, 227, 240, 250, 251, 288, 290, 292, 293, 307
Inter-American Convention on the Prevention, Punishment and Eradication of Violence Against Women, *see* Belém do Pará Convention
Inter-American Court of Human Rights (IACtHR), 20, 23, 24, 31, 34, 37–39, 41–43, 98, 105, 106, 113, 119, 123, 126, 127, 143, 204, 206–208, 212–214, 216, 218–220, 222, 225–227, 250, 251, 262, 263, 275
Intermediaries
defamation, 266
definition, 151
ECtHR, 172–175
EU law, 159–167
harassment, 246
hate speech, 302
human rights duties, 152
pornography, 334
self-regulation, 155
sexual violence, 228–230
International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), 7, 93, 94, 110, 164, 191, 287, 288, 290, 291, 293, 299, 303, 310, 352
International Covenant on Civil and Political Rights (ICCPR), 7, 21, 33, 109, 110, 112, 125, 127, 182, 257, 262, 273, 288–291, 293, 299, 303, 307, 310
International Covenant on Economic, Social and Cultural Rights (ICESCR), 2, 21, 28, 238
International Criminal Court (ICC), 207, 219
International criminal law (ICL), 12, 207, 208, 218–222, 232, 233
International Criminal Tribunal for Rwanda (ICTR), 111, 207, 208, 221, 222
International Criminal Tribunal for the Former Yugoslavia (ICTY), 111, 208, 219–221, 251
International Labour Organization (ILO), 239, 242
Internet Service Provider (ISP), 16
Intersectionality, 9, 52, 306, 312
ISP, *see* Internet Service Provider (ISP)
Istanbul Convention, 9, 36–39, 41, 49, 93, 98, 155, 158, 165, 205–207, 209, 210, 216, 218, 220, 221, 226, 227, 235, 237, 240–243, 252, 308, 329

L
Lanzarote Convention, 11, 19, 43, 111, 155, 209, 222, 317, 318, 323, 325, 328
LGBTQ, 204, 253, 292, 296
Liberalism, 23, 82, 87, 102, 129, 334
Linguistic philosophy, 10, 86, 88, 95, 100, 241

M
Maputo Protocol, 36–38, 41, 45, 98, 205–207, 238, 241, 308, 332
Margin of appreciation, 67, 107, 110, 120, 139, 141, 148, 150, 162, 173, 175, 191, 193, 216, 219, 222, 266, 267, 271, 295, 298, 316, 325, 327, 328, 335, 350–352

Marketplace of ideas, 17, 84, 102, 146, 185, 286, 318
Media publishers
liability, 167–172

N
Non-discrimination
equality, 21
gender-based violence, 36

O
OAS, *see* Organization of American States (OAS)
Objectification, 9, 37, 41, 52, 69, 78, 90, 97, 282, 304, 318, 320, 322, 344
Obscenity, *see* Pornography
OECD, *see* Organization for Economic Co-operation and Development (OECD)
Organization for Economic Co-operation and Development (OECD), 3, 29, 152
Organization of American States (OAS), 28

P
Pornography
distribution, 325–331
equality, 317, 331–334
harm, 315
international human rights law, 324–334
internet, 321–324
Postmodern feminism, 21
Privacy
autonomy, 123
disclosure, 131
harassment, 240
internet, 128–136
reasonable expectation, 130
rights theories, 122–127
sexual violence, 135, 206
Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa, *see* Maputo Protocol
Public/private divide, 82, 129–131, 137, 141, 143, 263, 270, 283, 345, 346, 349

R
Radical feminism, 10, 23, 63
Rape
definition, 218
gender, 206
human rights violation, 212
by proxy, 221, 223
Reputation, right to, 128
Revenge pornography, *see* Image-based sexual abuse
Right to be forgotten, 138, 174, 183, 185, 229, 280, 353

S
Sexism, 42, 44, 61, 69, 86, 87, 90, 91, 303–306, 309, 335
Sexist hate speech, 303
Sextortion, 48, 77, 135, 209–211, 222, 224, 232, 234, 254, 272, 349
Sexual harassment
definition, 236
discrimination, 241
domestic law, 237
international human rights law, 237
internet, 243–248
Sexual violence
criminalisation, 216
definition, 207
hierarchy, 212
intermediary liability, 228
investigation, 225
prohibition, 205
Spam, 82, 124, 246, 247, 249, 330, 335, 354
Stereotypes
online, 52
Stereotyping
international human rights law, 40–45
pornography, 320
sexist hate speech, 308

T
Threats, 250–255
Torture
sexual violence, 212
threats, 250

U
UDHR, *see* Universal Declaration of Human Rights (UDHR)
UNCAT, *see* UN Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment
UN Committee against Torture, 206
UN Committee on Economic, Social and Cultural Rights, 205

UN Committee on the Elimination of All Forms of Discrimination against Women, 30, 31, 33, 37–39, 42–44, 47, 49, 97, 100, 135, 205, 207, 213, 216–220, 223, 225, 227, 230, 238, 241–243, 247, 252, 300, 307, 319, 321, 331, 333, 335, 348, 352
UN Committee on the Elimination of Racial Discrimination, 31, 106, 109, 121, 155, 289, 291, 293, 298–300, 306
UN Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, 206
UN Convention on the Elimination of All Forms of Discrimination against Women, 41
UN Convention on the Rights of Persons with Disabilities, 25
UN Convention on the Rights of the Child, 209, 328
UNHRC, *see* UN Human Rights Committee (UNHRC)
UN Human Rights Committee (UNHRC), 33, 93, 97, 112, 205, 298, 332
Universal Declaration of Human Rights (UDHR), 112, 125
UN Special Rapporteur on the Freedom of Expression, 1, 27, 29, 33, 34, 47, 99, 106, 108, 113, 116, 153, 155, 156, 182, 189, 256, 262, 266, 293, 294, 298, 300, 301, 307, 309, 310, 329
UN Special Rapporteur on Torture, 37, 207, 213
UN Special Rapporteur on Violence against Women, 4, 30, 35, 40, 47, 48, 52, 77, 98, 135, 145, 185, 216, 238, 243, 253, 254, 256, 308, 333, 335

V

Vulnerable group, 11, 110, 150, 310