# FREEDOM OF EXPRESSION ON SOCIAL MEDIA –

# CONCEPTUAL AND REGULATORY CHALLENGES IN EUROPE AND THE US

## A. INTRODUCTION

Social media platforms have revolutionised our ability to connect across historic social, political and geographic divides. Where previously gatekeepers mitigated and negotiated access to mass media platforms, today potentially anyone – and any content – can reach millions of individuals in an instant. This development bears great opportunities for the democratisation of expression and the diversification of public discourse but has likewise broadened the impact and harm done through disinformation and hate speech.

At first blush, there are marked differences between the exercise of freedom of expression in offline versus online scenarios. For instance, whereas hateful rhetoric in an offline environment may reasonably attract police intervention, arrest and criminal sanctions,[1] posting similar content on social media is less likely to attract comparable repercussions.[2] This discrepancy arises in part because it is unclear whether all forms of harmful speech unequivocally violate the terms and conditions (also known as community standards) of social media platforms in theory, and certainly in practice, companies like Facebook and twitter have been reluctant to fully enforce them. Even if content is found to violate community standards, the responses are typically limited to the deletion of the offending post and the potential suspension of user accounts following repeat violations.

All the while, criminal sanctions for online behaviour are rare due to the significant investigatory challenges that relative online anonymity presents, limited policing resources and underreporting of potential offenses. These differences in treatment and legal consequences raise the question whether there is something novel about social media that should occasion us to re-evaluate the principles underlying freedom of expression. How can these platforms be regulated effectively while recognising the importance of free and open discourse? The paper

[1] 'Racist chants at Nottingham Trent University: Two men released' *BBC News* (9 March 2018) <http://www.bbc.co.uk/news/uk-england-nottinghamshire-43342058> accessed 9 March 2018.
[2] 'YouTube attacked over Neo-Nazi National Action video' *BBC News* (7 March 2018) <http://www.bbc.co.uk/news/technology-43319975> accessed 9 March 2018; 'My online stalker: The feeling he could hurt me never went away' *BBC News* (9 March 2018) <http://www.bbc.co.uk/news/uk-england-hereford-worcester-43291038> accessed 9 March 2018.

does not propose to give a final answer to any of these questions. Rather is an attempt to guide and inform this important debate as lawmakers consider regulatory measures. Specifically, the paper investigates three pertinent aspects of the debate: the controversial German attempt to regulate expression on social media and the conventionally accepted legal principles of free expression under the ECHR and the First Amendment to the US Constitution.

Section B considers the regulatory response to social media platforms in Germany, which prominently and controversially introduced the so-called Network Enforcement Law (*Netzwerkdurchsetzungsgesetz* - NetzDG). The section examines how the regulation operates and evaluate the law based on relevant principles from the German Constitutional Court on freedom of expression. It will be argued that the core question for the constitutionality of the law is whether fears of 'overblocking' are borne out by the evidence. The paper argues that NetzDG potentially strikes an appropriate balance between open and free expression and the legitimate interest in asserting well-established offline limitations in online environments. This regulatory experiment in a major European jurisdiction has significant practical relevance as regulators across Europe and the United States weigh responses to the legal novelties and challenges that social media represents.

Section C then examines the legal environment in Europe and the US on freedom of expression, highlighting the areas of particular relevance to the regulation of social media. Under the ECHR, state limitations on freedom of expression are generally viewed as an acceptable means of regulating public discourse and safeguarding competing rights of individuals. This generally permits limitations for the purposes of combatting hate speech. However, the Court equally recognises the chilling effect of criminal sanctions, and the fact that a liberal democratic state has an inherent interest in free expression and platforms that permit its effective exercise. This of course does not necessarily equate to sanctioning any the expression of any opinion short of hate speech, nor granting individuals access to any platform they desire: reasonable limitations arise from competing rights. However, the Court is notably open to state measures that seek to guarantee freedom of expression of individuals against encroachments from private actors.

This stands in sharp contrast to the case law of the US Supreme Court under the Frist Amendment to the US Constitution. The Court is generally sceptical towards any government limitations on speech and follows a doctrine of strict content neutrality. The latter generally renders unconstitutional any laws designed to combat hate speech, as these invariably sanction the expression of certain opinions, but not others. Save for narrow exceptions (e.g. incitements to violence, fighting words and obscenity), limitations are generally not permissible under this

broad interpretation. Given this strong stance, it is perhaps somewhat surprising that the Supreme Court has almost nothing to say about private entities limiting freedom of expression. Private entities are generally free to impose any restrictions they see fit, as the First Amendment is understood to exclusively apply to government action. While there were some hints that the Supreme Court might extend protections to private actions, these are isolated rulings that currently do not form a coherent exception.

The paper concludes that given these differences in legal environments, it is unlikely that the regulatory approach of NetzDG can be easily generalised. While the approach may well be compatible with the German Constitution and the ECHR, it is very unlikely that the law would survive scrutiny under the US First Amendment. This necessitates not only distinct regulatory frameworks for each jurisdiction, but also goes some way to explain the different approaches in political activism and lobbying. Paradoxically, the community standards developed by Facebook and twitter are generally acceptable in both jurisdictions, albeit for very different legal reasons and subject to some caveats.

## B. NETZWERKDURCHSETZUNGSGESETZ

The German Network Enforcement Law (NetzDG)[3] has attracted much media attention since fully entering into force on 1 January 2018.[4] It was sparked in part due to a number of high profile deletions of content from German politicians, including a tweet from Heiko Maas, then Minister for Justice and primarily responsible for the legislation.[5]

This section will give a succinct overview of the NetzDG, clarifying some common misconceptions in media and scholarly treatments of the legislation. It will then evaluate the

[3] *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz)*, Act to Improve Enforcement of the Law in Social Networks of 1 September 2017 (BGBl. I 3352); an English translation can be found on the website of the Federal Ministry of Justice: 'Act to Improve Enforcement of the Law in Social Networks' (*Bundesministerium der Justiz und für Verbraucherschutz*, 12 Juli 2017) <http://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf;jsessionid=E23D49344654C18787B25C74A2E3D9C9.1_cid334?__blob=publicationFile&v=2> accessed 1 July 2018.

[4] See for instance 'Companies face 50m euro fines in Germany for hate speech' *BBC News* (3 October 2017) <https://www.bbc.co.uk/news/technology-41483412> accessed 1 July 2018; 'Tough new German law puts tech firms and free speech in spotlight ' *Guardian* (5 January 2018) <https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight> accessed 1 July 2018.

[5] Andrea Diener, 'Storch-Satire ist nicht regelkonform' *Frankfurter Allgemeine Zeitung* (3 January 2018) <http://www.faz.net/aktuell/feuilleton/medien/twitter-sperrt-titanic-magazin-wegen-storch-satire-15371919.html> accessed 1 July 2018; 'Maas-Tweet über Thilo Sarrazin gelöscht' *Spiegel Online* (8 January 2018) <http://www.spiegel.de/netzwelt/netzpolitik/netzdg-heiko-maas-tweet-ueber-thilo-sarrazin-verschwunden-a-1186747.html> accessed 1 July 2018.

provisions against the backdrop of the principles identified by the ECtHR and conclude that some common criticisms are overstated. The NetzDG is unlikely to resolve all challenges surrounding social media and freedom of expression, and undoubtedly risks stifling expression online. However, apart from this risk, it also promises to contribute to more inclusive debates by giving loud and radical voices less prominence. In any case, it appears reasonable to let this regulatory experiment play out and observe whether fears over a 'chilling effect' are borne out by the evidence. A review of the law and its effects are is planned after an initial three-year operation period, which should deliver ample data and regulatory experience while limiting the scope for potential harm.[6]

### *1. The statute in a nutshell*

The NetzDG applies to 'social media networks' operators with at least two million users within Germany. Social media networks are defined as internet platforms that seek to profit from providing users with the opportunity to share content with other users and the broader public. Platforms which provide individualised communication services, such as email or messaging apps, as well as platforms providing editorialised content, such as news websites, are explicitly excluded from the scope of the law (§ 1 I NetzDG).

The core obligations centre on establishing an effective and transparent complaints management infrastructure (§ 3 NetzDG) and compiling bi-annual reports on the complaints management activity (§ 2 NetzDG). Especially the latter reporting obligations are quite detailed and include provisions that set out training and management oversight requirements. The complaints management infrastructure must chiefly ensure that the social networks delete or block 'illegal content' within a specified timeframe. Deletion results in a global removal of the content from the platform, while blocking merely makes the content unavailable in Germany. Although blocking and deleting are thus distinct, they will be collectively referred to as 'deleting' throughout this section.

Content is designated illegal if it falls under the one of the enumerated provisions of the German criminal code (StGB).[7] The most important ones for the purposes of freedom of expression are insult (§ 185), defamation (§ 186 and § 187), public incitement to crime (§ 111), incitement to hatred (§ 130) and dissemination of depictions of violence (§ 131). In this context, it is

[6] *Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG)* (Drucksache 18/12356, 2017) <http://dipbt.bundestag.de/dip21/btd/18/123/1812356.pdf#page=18> accessed 1 July 2018.

[7] Criminal Code in the version promulgated on 13 November 1998 (BGBl. I 3322), last amended by Article 1 of the Law of 24 September 2013 (BGBl. I 3671) and Article 6(18) of the Law of 10 October 2013 (BGBl. I 3799).

important to note that the obligation to delete is not novel. The NetzDG merely enforces compliance with existing legal obligation under § 10 of the Telemedia Act (TMG).[8] Under that provision, social media operators are liable for illegal content on their site under criminal and private law. The two core regulatory innovations arising from NetzDG are the deadlines for actioning reported content, and the possibility of fines for shortcomings of the complaints management procedure.

With respect to the deadlines, NetzDG distinguishes between 'manifestly illegal' and 'illegal' content, prescribing different deadlines for deletion. Manifestly illegal content must be deleted within 24 hours of a receiving a complaint, while merely illegal content must be actioned within a deadline of seven days. Manifestly illegal content is not defined more clearly in the law, but legislative commentary during the parliamentary process suggests that it would apply only to the most obvious of cases: if any doubt arises, the seven-day deadline would apply.[9]

The most important exception to the seven-day deadline applies if operators refer the decision of whether to delete to an independent body of industry self-regulation (§ 3 (3) (b) NetzDG).

Such bodies must be setup and funded collectively by social media platforms and ensure independent decisions that the operator accepts as binding. Certain conditions of operation apply to bodies of industrial self-regulation (§ 3 (6) NetzDG), and they must in any case be accredited by the Ministry of Justice before commencing operation. Such bodies of industry self-regulation are a common feature in the German regulatory landscape and have been setup for instance by the movie, tv, and video games industries to rate the age appropriateness of content.[10] Decisions from these bodies determine for instance the age restrictions for movies and video games, which have direct legal consequences affecting sales and marketing rules under youth and consumer protection laws. Practically speaking, industry self-regulation bodies hold considerable power and influence. For instance, the long-standing ban on any Nazi symbols in video games was largely based on the refusal of the competent self-regulation body USK refusing age certification unless such content was removed. It is important to note that this was not necessarily legally required under German law, as the relevant provisions contain

[8] Telemedia Act of 26 February 2007 (BGGl. I 179), last amended by Article 1 of the Act of 28 September 2017 (BGBl. I 3530).
[9] *Gesetzentwurf Netzwerkdurchsetzungsgesetz*.
[10] 'Freiwillige Selbstkontrolle der Filmwirtschaft - FSK' (*Freiwillige Selbstkontrolle der Filmwirtschaft GmbH*) <https://www.spio-fsk.de/> accessed 1 July 2018 [self-regulation body of the film industry]; 'Freiwillige Selbstkontrolle Fernsehen - FSF' (*Freiwillige Selbstkontrolle Fernsehen e.V.*) <https://fsf.de/> accessed 1 July 2018 [self-regulation body of the television industry]; 'Unterhaltungssoftware Selbstkontrolle - USK' (*Freiwillige Selbstkontrolle Unterhaltungssoftware GmbH*) <http://www.usk.de/> accessed 1 July 2018 [self-regulation body of the video games industry].

express exceptions for the use of these symbols for historic, scientific and artistic purposes (§ 86 (3) and 86a (3) of the German Criminal Code). USK has only recently announced a change in its evaluation process, moving towards a case-by-case analysis of the use of Nazi symbols, and thereby loosening its previous blanket ban.[11]

Separately from this duty to delete content, criminal prosecution remains possible for any content that violates the criminal code, and social media platforms must preserve deleted content for evidence purposes for ten weeks (§ 3 II 4 NetzDG). Additionally, NetzDG requires social media platform operators to name an agent in Germany that is responsible for receiving complaints. A failure to name or lack of response from a responsible agent attracts a fine of up to 500.000 Euros, while any other failure to implement a NetzDG compliant management scheme can draw a fine of up to 5 million Euros. The latter increases to 50 million Euros for legal persons and corporations under § 30 II 3 of the Act on Regulatory Offences, see § 4 II 2 NetzDG.[12]

### *2. An assault on freedom of expression?*

It is rare for a legal system to treat freedom of expression as an absolute right and as a matter of German constitutional law, it is not clear whether the provision would run afoul of freedom of expression. Initially, it is useful to distinguish two scenarios. In the first scenario, a social media platform operator deletes illegal content, while in the second scenario the operator mistakenly deletes legal content.

Under the German Basic Law, freedom of expression does not cover insults and defamations, or incitement to hatred.[13] To the extent that deletion of the illegal content amounts to an infringement, this is justified as it is provided by provisions of general (read: content neutral) laws under Article 5 II Basic Law.[14] Moreover, deleting illegal content appears as a measured sanction, given that such statements, when made in offline scenarios, often attract criminal prosecution which may result in fines and prison sentences.

With the second scenario, the issues become more complicated. The German Federal Constitutional Court has recognised that there is a presumption in favour of freedom of

[11] 'USK berücksichtigt bei Altersfreigabe von Spielen künftig Sozialadäquanz' (Berlin, 9 August 2018) <http://www.usk.de/service/presse/details-zum-presseartikel/article/usk-beruecksichtigt-bei-altersfreigabe-von-spielen-kuenftig-sozialadaequanz/> accessed 1 August 2018.
[12] Act on Regulatory Offences of 19 February 1987 (BGGl. I 602), last amended by Article 4 of the Act of 13 May 2015 (BGBl. I 706).
[13] Federal Constiutional Court, 1 BvR 400/51, 15 January 1958, BVerfGE 7,198; Federal Constitutional Court, 1 BvR 2150/08, 4 November 2009, BVerfGE 124, 300.
[14] There is one important exception to this rule that we discuss below.

expression whenever it is unclear whether the expression is illegal, at least on topics of public interest.[15] This has been settled case law ever since the famous decisions in Lüth and more recently in Wunsiedel.[16] Notably, the Court extends this protection to public forums, even where access to them is regulated through private law relationships. The relevant decision dealt with the question whether the right to protest extended to the publicly accessible areas of Frankfurt Airport, run by a corporation under a controlling majority of local, state and federal government.[17] The Federal Constitutional Court held that the corporation was directly bound to uphold constitutional rights just as any other state entity. The fact that it was incorporated under private law and managed its legal relationship with visitors through ordinary private law (as opposed to public law) was immaterial to this conclusion.[18] Protesters are free to choose the time, place and manner of the protest, provided that the chosen place is generally open to the public and not open merely for specific purposes, such as a public swimming pool or a hospital.[19] In particular, a wish to avoid inconveniences and spare visitors confrontation with unpleasant opinions were not deemed legitimate grounds for limiting freedom of expression.[20]

There was no need for the Court in this case to decide to what extent freedom of expression also imposed obligations on wholly private actors. However, it nonetheless mentioned the so-called *mittelbare Grundrechtsbindung*, that is to say the indirect obligations of private actors and *Schutzpflichten*, the positive obligations on the state from constitutional rights as avenues for such obligations.[21] In this context, the Court clarified that while these indirect obligations operated on a relative scale, as they ultimately mediate relationships between private individuals that are equally entitled to freedom of action and constitutional rights. However, this does not invariably mean, according to the Court, that the weight of the obligations would be lower on the private actors when compared to the state. Where private actors provide the infrastructure and forums necessary for individuals to communicate their ideas, a role previously at least partially assumed by the state (e.g. state monopolies on postal services and telecommunications), the obligations might well be indistinguishable.[22]

[15] Mathias Hong, 'The German Network Enforcement Act and the Presumption in Favour of Freedom of Speech' (*Verfassungsblog*, 22 January 2018) accessed 1 July 2018
[16] Federal Constiutional Court, BVerfGE 7,198; Federal Constitutional Court, BVerfGE 124, 300.
[17] Federal Constitutional Court, 1 BvR 699/06, 22 February 2011, BVerfGE 128, 226.
[18] Ibid at [46-48] and [53].
[19] Ibid at [65]; Federal Constitutional Court, 1 BvR 233/81, 1 BvR 341/81, 14 May 1985, BVerfGE 69, 315.
[20] Federal Constitutional Court, 1 BvR 1762/95, 1 BvR 1787/95, 12 December 2000, BVerfGE 102, 347; Federal Constitutional Court, BVerfGE 128, 226 at [103].
[21] Both are based on settled constitutional interpretation:
[22] Federal Constitutional Court, BVerfGE 128, 226 at [59].

NetzDG notably does not require censorship (i.e. pre-emptively scrutinizing content before it is shared, which is unconstitutional under Article 5 I 3 Basic Law),[23] nor does is it directed against any specific content, which, as mentioned previously would violate the content neutrality requirement and would generally be unconstitutional.[24] The only exception to this is the reference in NetzDG to § 130 (3) and (4) of the Criminal Code, which specifically criminalises approving of, glorifying, downplaying or justifying National Socialist rule and holocaust denial. These are evidently not content neutral provisions but were nonetheless expressly found constitutional in the *Wunsiedel* decision.[25] In a more recent decision, the Constitutional Court extended its reasoning for the constitutionality of § 130 (4) StGB to the similarly worded § 130 (3) StGB.[26]

Hence, the obligation to delete only stretches to content already illegal under constitutionally sanctioned criminal law provisions and primarily enforces obligations that already existed under § 10 TMG. The only real legal novelty provided by NetzDG is the introduction of deadlines for the processing of content reported to the platforms.

Hence, arguments alleging unconstitutionality of NetzDG cannot easily rely on its lack of content neutrality or the obligation to delete illegal content as such. Critics therefore often contend that NetzDG will promote an overly aggressive deletion policy of *legal* content (so-called 'overblocking') that will have a 'chilling effect' on freedom of expression for users of social media platforms: reducing their readiness to express themselves online. If overblocking does take place as a result of NetzDG, then this would indeed be would be problematic under the German Basic Law.

### *3. The overstated danger of overblocking*

Despite their prevalence in legal writing on the subject, concerns that social media platforms will delete content rather than risk a fine, appear overstated. Overblocking is likely to arise, so goes the argument, due to the structure of the fines that apply to a systematic failure to delete

[23] This is settled case law: Federal Constiutional Court, BVerfGE 7,198; Federal Constiutional Court, 1 BvR 657/68, 26 May 1970, BVerfGE 28, 282; Federal Constitutional Court, 1 BvR 934/82, 19 November 1985, BVerfGE 71, 162; Federal Constitutional Court, 1 BvR 1476/91, 1 BvR 1980/91, 1 BvR 102/92, 1 BvR 221/92, 10 October 1995, BVerfGE 93, 266.

[24] Federal Constitutional Court, BVerfGE 124, 300.

[25] § 130 of the German Criminal Code was deemed a justified infringement of freedom of expression given the historical context of the Basic Law as an antithesis to Nazi ideology, see ibid at [52].

[26] Federal Constiutional Court, 1 BvR 2083/15 22 June 2018, unpublished; Federal Constiutional Court, 1 BvR 673/18, 22 June 2018, unpublished; see critical commentary of Mathias Hong, 'Holocaust, Meinungsfreiheit und Sonderrechtsverbot – BVerfG erklärt § 130 III StGB für verfassungsgemäß' (*Verfassungsblog*, 5 August 2018) <https://verfassungsblog.de/holocaust-meinungsfreiheit-und-sonderrechtsverbot-bverfg-erklaert-%C2%A7-130-iii-stgb-fuer-verfassungsgemaess/> accessed 5 August 2018.

illegal content. Hence, a prudent social media platform operator would, when in doubt and confronted with a high volume of complaints, delete content that is questionable, rather than risk a fine.[27]

With respect to illegal content, the matter is unproblematic from a constitutional perspective. For the reasons outlined earlier, users do not benefit from protections under freedom of expression for illegal content. Again, the more problematic scenario arises when the social media platform mistakenly deletes *legal* content. For the user, this represents an infringement of freedom of expression. Indeed, if overblocking is a prevalent phenomenon beyond the occasional erroneous decision it could dissuade users from expressing their views on the platform. This in turn, would render the NetzDG significantly more problematic, and arguably unconstitutional.

The Federal Constitutional Court has found a violation in ordering the publishers of a satirical magazine to pay compensation to an individual for an allegedly defamatory article, chiefly basing their ruling on the risk that it would discourage future exercise of freedom of expression.[28] However, it is not clear that such a chilling effect is inevitable under NetzDG. Occasional, non-systematic mistakes by social media platforms in an otherwise lawful complaints management infrastructure would arguably not suffice to produce such an effect. Notably, and contrary to the impression given by some reports, no fines attach to decisions in individual cases.[29] Rather, fines require a systemic and persistent failure in the complaints management infrastructure which must be substantiated through content that has been ruled illegal by a court in a separate decision (§ 4 V NetzDG).

It is difficult to see why a social media platform, which ultimately requires continuous user engagement and content creation to be profitable would adopt an overly aggressive deletion policy. It would risk an exodus of users through a consistent deletion of legal content, and thus critically undermine the viability of the platform. It appears more likely that the limited scope

---

[27] Lukas Gerhardinger, 'Das Netzwerkdurchsetzungsgesetz: Im Zweifel gegen die Meinungsfreiheit?' (*Verfassungsblog*, 17 April 2017) <https://verfassungsblog.de/das-geplante-netzwerkdurchsetzungsgesetz-im-zweifel-gegen-die-meinungsfreiheit/> accessed 1 July 2018; Mathias Hong, 'Das NetzDG und die Vermutung für die Freiheit der Rede' (*Verfassungsblog*, 9 January 2018) <https://verfassungsblog.de/das-netzdg-und-die-vermutung-fuer-die-freiheit-der-rede/> accessed 1 July 2018; Diana Lee, 'Germany's NetzDG and the Threat to Online Free Speech' (*Case Disclosed - Media Freedom and Information Access Clinic Blog*, 10 October 2017) <https://law.yale.edu/mfia/case-disclosed/germanys-netzdg-and-threat-online-free-speech> accessed 1 July 2018.

[28] Federal Constitutional Court, 1 BvR 514/90, 25 March 1992, BVerfGE 86, 1 at [35].

[29] 'Germany implements new internet hate speech crackdown' *Deutsche Welle* (1 January 2018) <https://www.dw.com/en/germany-implements-new-internet-hate-speech-crackdown/a-41991590> accessed 1 July 2018; Mark Scott and Janosch Delecker, 'Free speech vs. censorship in Germany' *Politico* (4 Janaury 2018) <https://www.politico.eu/article/germany-hate-speech-netzdg-facebook-youtube-google-twitter-free-speech/> accessed 1 July 2018.

of the fines and the inherent economic interests of social networks will encourage a more nuanced policy: one that complies with existing laws but avoids removing more content than necessary.[30] This line of reasoning is supported by data published in the first bi-annual reports from social media platforms on their complaints processing. The reports reveal that both facebook and twitter have elected not to delete the vast majority of content reported to them under a seperate NetzDG procedure. This is problematic for critics as a cornerstone of the case for unconstitutionality of the law relies on demonstrating that overblocking is more than a theoretical possibility. However, even assuming a measurable 'chilling effect' this would not necessarily equate to the unconstitutionality of NetzDG.

### *4. The free speech environment of social media*

It is unclear whether freedom of expression necessarily equates to a right to access to all platforms of expression. The case law is limited on this legal question, but there are good reasons to suspect that freedom of expression has certain limitations in this respect. In an older ruling, the Federal Administrative Court declined to award a recipient of social security the necessary transportation costs to attend a protest. It found that freedom of expression could not be extended to requiring the state to provide an individual with the means to engage in any form of expression she deems necessary.[31]

Moreover, it is important to recall that access to and expression on most social networks is already significantly limited through private law terms and conditions. These grant platform operators wide-ranging powers to delete content or even indefinitely suspend accounts of users for actions that are unlikely to fall afoul of German criminal law. Against this backdrop, it is difficult to sustain an argument that the potential for unintended side effects of NetzDG would on its own suffice to find it unconstitutional. Social media platforms are hardly a free speech paradise: users already operate in an environment where arbitrary limitations are placed on freedom of expression, where deletion and suspension of accounts may occur without the possibility of appeal and where the terms and conditions of participation can be altered at the sole discretion of the platform operators at any time.

---

[30] There is an unfortunate plethora of examples where social media have opted not to delete content, at times with thereby stoking violence and lynching, see Michael Safi, 'Sri Lanka accuses Facebook over hate speech after deadly riots ' *Guardian* (14 March 2018) <https://www.theguardian.com/world/2018/mar/14/facebook-accused-by-sri-lanka-of-failing-to-control-hate-speech> accessed 1 August 2018; Shaikh Azizur Rahman, ''Fake news often goes viral': WhatsApp ads warn India after mob lynchings ' *Guardian* (13 July 2018) <https://www.theguardian.com/world/2018/jul/13/fake-news-whatsapp-ads-india-mob-lynchings> accessed 1 August 2018.

[31] Federal Administrative Court, 5 C 11383, 13 September 1985, BVerwGE 72, 113 at [17].

## *5. Conclusion*

Overall, NetzDG may be a civilizing influence on online debate, instead of having a one-sided chilling effect on freedom of expression. The fact that to date social media and online interaction more generally has created a space for a significantly more *laissez faire* approach to expression is neither here nor there on the question of constitutionality.

The obligations to delete illegal content are based on well-establish limits to freedom of expression, to which NetzDG chiefly adds a more robust enforcement mechanism. The real novely of the law is to introduce strict time limits and specify requirements for the complaints management system. The constitutionality of this regulatory approach under German law will depend to a considerable extent on an evaluation of the complaints management infrastructure. Should it consistently lead to a chilling effect on freedom of expression, then the argument for unconstitutionality grows stronger, but in my view by no means straight-forward. Conversely, the Constitutional Court would be less likely to take issue with this novel regulatory approach if the deletion of legal content is limited to individual, non-systematic mistakes.

The first round of bi-annual report from social media platforms certainly do not answer all of these questions. At least for the time being, the data does not support the idea of widespread overblocking. Certainly, more data and close scrutiny of the law remains essential, as is an open mind towards reforms if it proves deficient. There are sound historical and legal arguments that caution us against overly strong interference with freedom of expression, especially as it lies at the core of the democratic model of government. However, NetzDG nonetheless appears as a solid attempt to impose a certain baseline of limitations on freedom of expression in social media environments. It can perhaps serve as a starting point for similar regulatory efforts now under consideration in Europe and the US, provided that it is generally compatible with the respective legal frameworks. The paper will now offer an initial assessment of the underlining principles in these jurisdictions based on the ECHR and the case law of the US Supreme Court on the First Amendment to the US Constitution.

# C. FREEDOM OF EXPRESSION IN EUROPE AND THE US

## *I. Europe*

### *1. General overview*

The right to freedom of expression is contained in Article 10 ECHR. The Court has described it as an 'essential foundation of a democratic society' and a 'basic condition for the development of every man.'[32] The provision especially protects the expression of unpopular, controversial, shocking and offensive views, which flow from the pluralism, tolerance and openness that is at the core of the Convention system.[33]

Article 10 ECHR affords broad protections to virtually any expression. A key conceptual distinction is drawn between the expression of opinions and factual claims, although this only affects the justification stage. At the initial infringement stage, protections are extended to both opinions and factual claims, regardless of their accuracy. Expressions must not necessarily be relate to issues of public interest,[34] and protections extend to the form and appearance of the expression, which includes exaggerations and provocations.[35]

Important exceptions from this broad scope fall under the general category of hate speech, especially some forms of anti-Semitic and extremist expressions, as well as incitements to violence: these are often excluded entirely from the scope of Article 10 pursuant to Article 17 ECHR (Prohibition of abuse of rights).[36] The Court routinely denies protection to holocaust denialism and calls for racial purity of states as contrary to historical fact and the values of the Convention.[37] In the words of the Court the Article was designed to '(…) make it impossible

---

[32] *Handyside v United Kingdom*, 5493/72, Plenary, 7 December 1976 at [49].
[33] *Sunday Times v United Kingdom*, 6538/74, Plenary, 26 April 1979 at [65]; Christian Mensching, 'Artikel 10 EMRK' in U Karpenstein and FC Meyer (eds), *Europäische Menschenrechtskonvention - Kommentar* (2nd edn, C.H. Beck 2015), p. 312-3.
[34] *Juppala v Finland*, 18620/03, Court, 2 December 2008 at [40] [abuse claims voiced in course of doctor's appointment]; *Jacubowski*, 15088/89, Court, 23 June 1994 at [25] [applicant instigated a public relations campaign to present his version of events regarding the circumstances of his dismissal]; *Kyprianou v Cyprus*, 73797/01, Grand Chamber, 15 December 2005 at [151] [lawyer pleading on behalf of client in court]; *Casado Coca v Spain*, 15450/89, Court, 24 February 1994 at [35] [commercial expression]; *Tatar and Faber v Hungary*, 26005/08 & 26160/08, Court, 12 June 2012 at [29] [non-verbal expression as protest].
[35] *Perna v Italy*, 48898/99, Grand Chamber, 6 May 2003 at [39]; *Mamère v France*, 12697/03, Court, 7 November 2006 at [25].
[36] *Hizb Ut-Tahrir and others v Germany*, 31098/08, Court, 12 June 2012 at [73] – [74]; *Ivanov v Russia*, 35222/04, Court, 20 February 2007 at [1]; *Norwood v United Kingdom*, 23131/03, Court, 16 November 2004.
[37] This is settled case law, on holocaust denial and some forms of antisemitism, see *Remer v Germany*, 25096/94, Commission, 6 September 1995; *Honsik v Austria*, 25062/94, Commission, 18 October 1995; *Lehideux and Isorni v France*, 24662/94, Grand Chamber, 23 September 1998 at [47]; *Garaudy v France*, 65831/01, Court, 24 June 2003; *Witzsch v Germany*, 7485/03, Court, 13 December 2005; *Ivanov v Russia*, 35222/04 at [1]; on calls for racial purity, see *Glimmerveen and Hagebbeek v Netherlands*, 8348/78; 8406/78, Commission, 11 October 1979.

for individuals to take advantage of a right with the aim of promoting ideas contrary to the text and the spirit of the Convention.'[38]

Incitements to violence constitute an illegitimate expression that may be banned and sanctioned in accordance with Article 10 ECHR. In the case of *Sürek v. Turkey*, the applicant published letters calling for violence against Turkish security forces and was found criminally liable by domestic courts. The Court ultimately declined to find a violation of Article 10 ECHR in light of the incitement to violence and the editorial responsibility for the publication of 'hate speech and the glorification of violence.'[39]

With respect to more general hate speech, the Court typically engages with the substance of the application, but draws on the rationale of Article 17 ECHR to justify infringements under Article 10 (2).[40] Crucially, hate speech does not require any incitement to violence, nor must it necessarily encourage the commission of crimes. This is an important distinction from the US Supreme Court approach under the so-called Brandenburg test, which we shall explore in more detail below. Suffice it to say here that justified limitations can be imposed on freedom of expression under the ECHR fir stirring up and justifying violence, hatred and intolerance.[41] This is especially the case where the expression denies the fundamental dignity and equality of human beings in a pluralistic, democratic society.[42]

### *2. Infringements*

#### *a. State*

Article 10 (2) names formalities, conditions, restrictions or penalties as examples of infringements of freedom of expression, although this list cannot be considered exhaustive.[43] There is a helpful distinction to be drawn between direct and indirect infringements.

Direct infringements are directed against the act of expression, by rendering it more difficult or impossible through measures designed to act as a prior restraint, or subsequent sanction. Court

---

[38] *Witzsch v Germany*, 7485/03 at [3].
[39] *Sürek v Turkey*, 26682/95, Grand Chamber, 8 July 1999 at [62].
[40] *Kühnen v Germany*, 12194/86, Commission, 12 May 1988; *B.H., M.W., H.P. and G.K. v Austria*, 12774/87, Commission, 12 October 1989; *Ochensberger v Austria*, 21318/93, Commission, 2 September 1994; *Nationaldemokratische Partei Deutschalnds, Bezirksverband München-Oberbayern v Germany*, 25992/94, Commission, 29 November 1995; *Rebhandl v Austria*, 24398/94, Commission, 16 January 1996; *Nachtmann v Austria*, 36773/97, Commission, 9 September 1998; *Schimanek v Austria*, 32307/96, Commission, 1 February 2000; *Soulas and others v France*, 15948/03, Court, 10 July 2008 at [42].
[41] *Feret v Belgium*, 15615/07, Court, 16 July 2009 at [73].
[42] *Perincek v Switzerland*, 27510/08, Grand Chamber, 15 October 2015 at [209] – [212]; *Feret v Belgium*, 15615/07 at [64]; *Gündüz v Turkey*, 35071/97, Court, 4 December 2003 at [40].
[43] Christian Mensching, 'Artikel 10 EMRK', p. 319.

orders preventing publications are a classic example of prior restraint,[44] while criminal or private law action are examples of subsequent sanctions.[45] The case of *Open Door and Dublin Well Woman v. Ireland* concerned an Irish Supreme Court injunction placed on the applicants, preventing them from counselling pregnant women on abortion facilities available outside of Ireland. The injunction itself was based primarily on the constitutional ban on abortion and to a lesser extent on statutory provisions that criminalised abortion and permitted the censorship of information material promoting abortion. The ECtHR took this opportunity to clarify that a law may violate the Convention rights even in the absence of a specific individual measure of implementation.[46] Ultimately, it concluded that the restraint imposed on the applicants was disproportionate to the aims pursued, and thus found a violation of Article 10 ECHR.[47] This was primarily based on the broad and perpetual nature of the injunction, especially given the fact that travel to a jurisdiction for the purposes of abortion as suggested in some of the material was not as such illegal in Ireland.

Indirect infringements occur when the exercise of freedom of expression is not expressly restrained or subsequently sanctioned, but nonetheless certain repercussions attach to the exercise of the right. To the extent that these sanctions are related to the expression and complicate or render impossible expression, these constitute infringements of Article 10 ECHR. This is especially relevant as such sanctions can produce a chilling effect for future expression. In the case of *Vogt v. Germany* the applicant was dismissed from her post as teacher, a civil service position, due to her political activities for the German Communist Party.[48] Despite the Court's previous rulings that had established that states were entitled to restrict and condition access to civil service roles based on political beliefs and loyalty to the constitutional order,[49] the Grand Chamber found no difficulty moving beyond the infringement stage. This is typical of freedom of expression cases, where the Court tends to take a broad view of infringements and typically affords the complicated issues deeper consideration at the justification stage. It was notable that the German authorities had been aware of the applicant's political activities before she was appointed a permanent civil servant and the party she was involved with was not banned by the Federal Constitutional Court. Hence, her actions on behalf of the party were

[44] *Observer and Guardian v United Kingdom*, 13585/88, Plenary, 26 November 1991.
[45] *Juppala v Finland*, 18620/03 at [40].
[46] *Open Door and Dublin Well Woman v Ireland*, 14234/88, Plenary, 29 October 1992 at [44].
[47] Ibid at [80].
[48] *Vogt v Germany*, 17851/91, Grand Chamber, 26 September 1995.
[49] *Kosiek v Germany*, 9704/82, Plenary, 28 August 1986; *Glasenapp v Germany*, 9228/80, Plenary, 28 August 1986.

entirely lawful. These were key considerations for the Court to find a violation of Article 10 ECHR in her dismissal.

In the similar case of *Wille v. Liechtenstein*, the former President of the Liechtenstein Administrative Court had complained that a pre-emptive announcement of refusal by the Prince of Liechtenstein to appoint him to public office in the event of his election constituted a violation of his freedom of expression.[50] The announcement had been preceded by an argument between the Prince and members of the government, which was ultimately settled through a joint declaration, involving the applicant. The applicant had expressed the view in public lectures that the Liechtenstein Constitutional Court may interpret the law to resolve disagreements between the Prince and the People. The Prince had understood this to draw into question the applicant's commitment to the Constitution of Liechtenstein, which states that the Prince and People are jointly sovereign. The sanction threated by the Prince was never enacted, as the Diet ultimately did not successfully elect the applicant to public office. The Court therefore had to consider whether the actions of the Prince amounted to an interference with Article 10 ECHR rights of the applicant. The Court concluded that an interference indeed arose from the announcement by the Prince.[51] The Court relied expressly on the chilling effect of the announcement on the applicant and his future willingness to exercise his freedom of expression with regard to constitutional matters.

Sanctions do not necessarily need to attach to the content of the expression, as Article 10 ECHR also covers the right of individuals to choose the form of expression they deem to be most effective. In the case of *Women on Waves v. Portugal*, the applicants were prevented by a navy warship from sailing in Portuguese territorial waters for the purposes of expressing their views on abortion.[52] The Court reiterated that Article 10 protects the content, as well as the means of expression. Hence, it found a violation in preventing the use of Portuguese waters.[53]

### *b. Private*

Beyond the requirement to refrain from unjustified infringements of Article 10 ECHR, states are also under an obligation to protect individuals from infringements through third parties.[54]

---

[50] *Wille v Liechtenstein*, 28396/95, Court, 28 October 1999 at [6] – [12].
[51] Ibid at [50].
[52] *Women on Waves and others v Portugal*, 31276/05, Court, 3 February 2009 at [15].
[53] Ibid at [30].
[54] *Palomo Sanchez and others v Spain*, 28955/06, 28957/06, 28959/06 and 28964/06, Grand Chamber, 12 September 2011 at [59]; *Özgür Gündem v Turkey*, 23144/93, Court, 16 March 2000 at [42] – [46]; *Fuentes Bobo v Spain*, 39293/98, Court, 29 February 2000 at [38]; *Dink v Turkey*, 2668/07, 6102/08, 30079/08, 7072/09, 7124/09, Court, 14 September 2010 at [106].

This positive obligation does however have some limits when compared to protection from state action.

For one, the protection does not amount to a duty to provide individuals with access to any and all forums for expression they choose: states are under no obligation to make the use of private or public property available to individuals, provided that the use is not indispensable to the exercise of the right to free expression. In the case of *Appleby*, the applicants sought to set up a stand and distribute leaflets in a privately-owned shopping mall.[55] The Court reiterated the positive obligations for states under Article 10 ECHR, which may require facilitating access to privately owned property. However, freedom of expression does not 'bestow any freedom of forum for the exercise of that rights.'[56] Rather, the denial of access must have the '(…) effect of preventing any effective exercise of freedom of expression or it can be said that the essence of the right has been destroyed (…)', for instance in the context of a corporate town, where the entire municipality is privately owned.[57] Crucially, the right also does not require access to the most convenient forum of expression. The crucial question for the Court was thus whether the applicants were effectively prevented from exercising their right to expression. Ultimately, the Court concluded that owing to the (however limited) access to available to the shopping centre and the existence of further alternative forums that the rights of the applicants were not violated.[58]

### *3. Justification*

Generally, infringements of Article 10 ECHR may be justified where they are prescribed by law, pursue a legitimate aim and are necessary in a democratic society. Considering the broad scope of freedom of expression, this is the crucial stage to determine the effective limits under the Convention system. The Court typically takes a restrictive view on justifications of infringements under Article 10 (2) ECHR, holding that exceptions must be 'narrowly interpreted and the necessity for any restrictions must be convincingly established.'[59] Infringements on freedom of expression require some basis in domestic law,[60] and the law must be '(…) accessible to the person concerned, who must moreover be able to foresee its

[55] *Appleby and others v United Kingdom*, 44306/98, Court, 6 May 2003 at [40] – [48].
[56] Ibid at [47].
[57] Ibid at [47].
[58] Ibid at [50].
[59] *Observer and Guardian v United Kingdom*, 13585/88 at [59].
[60] *Association Ekin v France*, 39288/98, Court, 17 July 2001 at [44].

consequences, and that it should be compatible with the rule of law.'[61] Legislation with excessively broad and unreviewable margins of appreciation for authorities is generally insufficient to meet this requirement.

At the justification stage the aforementioned distinction between opinion and factual claims becomes relevant. The core difference between the two is that factual claims may be at least theoretically proven or disproven, while opinions are not open to such proofs.[62] In the instructional case of *Lingens v. Austria*, the applicant faced criminal sanctions for allegedly defamatory statements directed against an Austrian politician based on his Nazi past. The Court upheld the distinction between the (uncontested) factual claims that the politician had served in the SS and the expression of views in two newspaper articles authored by the applicant. This was of central importance for the Court to find a violation of the Convention based on certain provision of the Austrian criminal code which required the applicant to prove the truthfulness of his opinion in order to avoid criminal sanctions. The Court held that conceptually, this was impossible given the nature of an opinion as fundamentally unprovable, and hence Austria was found to have violated Article 10 ECHR.[63]

The Court recognizes that one cannot always draw a clear distinction between factual claims and opinions. Hence it has accepted the existence of 'value laden statements of fact', which are ultimately treated as expressions of opinion. In the case of *Karsai v. Hungary*, the applicant was subjected to a fine for allegedly criticising the leader of a right wing political movement against the backdrop of a debate over Hungary's role in the Second World War and the Holocaust. The Court relied on the opinion elements in the statements made by the applicant to determine that it was not a simple statement of fact, and hence a proof of truthfulness was not required as assumed by the domestic courts.[64] Such cases understandably blur the distinction between factual claims and opinions, which has proved difficult to maintain in most Article 10 ECHR cases. The Court has adopted a nuanced view, holding that '(…) a value judgment must be based on sufficient facts in order to constitute a fair comment under Article 10 (…)' and that the core difference between opinions and factual claims '(…) finally lies in the degree of factual proof which has to be established.'[65] However, when in doubt, the Court tends to hold that the

[61] This is settled case law, see *Kruslin v France*, 11801/85, Court, 24 April 1990 at [27]; *Sunday Times v United Kingdom*, 6538/74 at [49]; *Lindon, Otchakovsky-Laurens and July v France*, 21279/02, 36448/02, Grand Chamber, 22 October 2007 at [41]; *Amihalachioaie v Moldova*, 60115/00, Court, 20 April 2004 at [25].
[62] *Lingens v Austria*, 9815/82, Plenary, 8 July 1986 at [46].
[63] Ibid at [46] and [47].
[64] *Karsai v Hungary*, 5380/07, Court, 1 December 2009 at [33] - [38].
[65] *Scharsach and News Verlagsgesellschaft mbH v Austria*, 39394/98, Court, 13 November 2003 at [40]

expression constitutes an opinion, and emphasises its autonomous determination from domestic court judgments.[66] Following the initial determination of whether an expression constitutes a factual statement, an opinion or a mixture of both, the Court attaches different justification requirements on infringements.

With respect to factual statements, applicants can be required to prove the veracity of their claims with respect to defamatory claims, and must in any case be given an opportunity to do so, as otherwise this in itself constitutes a violation of Article 10 ECHR.[67] However, even if a factual statement is proven untrue, it remains to be determined whether the applicant acted in good faith, with the standard of care being higher for more serious allegations.[68] This does not prevent limitations on the publication of true factual statements, to the extent that there is no public interest in their dissemination.[69]

With expressions of opinion, any obligation to prove the veracity is impossible to meet and as such constitutes a violation of Article 10 ECHR.[70] Nonetheless, the Court holds that even opinions must have a sufficient basis in fact.[71] The presence or absence of such a basis in fact is an important aspect in the overall proportionality analysis of any infringement and may render an opinion excessive.[72] However the Court has been sceptical towards applying a strict evidentiary standard with respect to the factual basis.[73] Strong and harsh opinions are permissible as long as they do not constitute a gratuitous personal attack and seek only to defame an individual. In the case of *Gorelishvili v. Georgia*, the Court found that while the statements made by a jounralist regarding a property deal involving a member of parliament were '[s]arcastic and cynical', they did not amount to '(…) a gratuitous personal attack upon

---

[66] *Roland Dumas v France*, 34875/07, Court, 15 July 2010 at [55]; *Wirtschafts-Trend Zeitschriften Verlagsgesellschaft M.B.H. (No 3) v Austria*, 66298/01 and 15653/02, Court, 13 December 2005 at [42] – [44].
[67] *Steel and Morris v United Kingdom*, 68416/01, Court, 15 February 2005 at [93]; *McVicar v United Kingdom*, 46311/99, Court, 7 May 2002 at [84]; *Colombani and others v France*, 51279/99, Court, 25 June 2002 at [66] [no opportunity given to prove veracity].
[68] *Rumyana Ivanova v Bulgaria*, 36207/03, Court, 14 February 2008 at [64]; *Ringier Axel Springer Slovakia, A.S. v Slovakia*, 41262/05, Court, 26 July 2011 at [97]; Christian Mensching, 'Artikel 10 EMRK', p. 329.
[69] *Ruusunen v Finland*, 73579/10, Court, 14 January 2014 at [51] and [52] [criminal sanctions for disclosure of private details of a romantic affair with Prime Minister in a book publication]; *Standard Verlags GMBH v Austria*, 21277/05, Court, 4 June 2009 at [52] [publication of gossip regarding the marriage of the Federal President]; *Éditions Plon v France*, 58148/00, Court, 18 May 2004 at [50].
[70] *Lingens v Austria*, 9815/82 at [46]; *OOO Ivpress and others v Russia*, 33501/04, 38608/04, 35258/05, 35618/05, Court, 22 January 2013 at [72].
[71] *Jerusalem v Austria*, 26958/95, Court, 27 February 2001 at [43]; *De Haas and Gijsels v Belgium*, 19983/92, Court, 24 February 1997 at [47].
[72] *Feldek v Slovakia*, 29032/95, Court, 12 July 2001 at [75] and [76]; *Steel and Morris v United Kingdom*, 68416/01 at [87].
[73] *Brosa v Germany*, 5709/09, Court, 17 April 2014 at [47] and [48].

the parliamentarian (…)' as they had some factual basis.[74] Especially public figures must generally endure more robust and polarising expressions of opinion,[75] but only to the extent that the expression is relevant to their public life. If the expression is merely intended to satisfy the curiosity of the public, stronger limitations are permissible.[76]

## *II. Free speech in the US*

### *1. General overview*

The scope of protection currently afforded from state interference under the First Amendment is exceptionally broad. The First Amendment generally convers speech without regard to whether it is of a political or commercial nature. It protects from most prior restraints, includes a right to refrain from speech and extends to most forms of hate speech. The US Supreme Court generally takes the view that state restrictions are unlawful, except for a set of narrowly prescribed circumstances. Among the more prominent exceptions are incitement to violence, fighting words, and obscenity. In contrast to its European counterpart, the Supreme Court recognises few positive obligations or obligations for private entities under the First Amendment. As a result, private entities are free to impose virtually any limitation on speech they desire. This is a gap in the otherwise comprehensive protection regime, which has significant implications for the regulation of expression on social media, as these are dominated by private entities.

### *2. Hate speech*

The Supreme Court follows a doctrine of strict content neutrality which prevents limitations designed to tackle hate speech. This was established in the case of *Smith v. Colin*.[77] The Nationalist Socialist Party of America group had planned a march in the town of Skokie in uniform with swastikas and banners, a Chicago suburb of with a significant Jewish population.

---

[74] *Gorelishvili v Georgia*, 12979/04, Court, 5 June 2007 at [40]; *Katamadze v Georgia*, 69857/01, Court, 5 June 2007.
[75] *EON v France*, 26118/10, Court, 14 March 2013 at [59] [alleged defamation of the French President]; *Mondragon v Spain*, 2034/07, Court, 15 March 2011 at [50] [alleged defamation of the Spanish King]; *July and SARL Liberation v France*, 20893/03, Court, 14 February 2008 [74]; *Dupuis and others v France*, 1914/02, Court, 7 June 2007 at [40]; *Steel and Morris v United Kingdom*, 68416/01 at [94]; *Jerusalem v Austria*, 26958/95 at [38].
[76] This is settled case law: *Axel Springer AG v Germany*, 39954/08, Grand Chamber, 7 February 2012 at [91]; *Saaristo and others v Finland*, 184/06, Court, 12 October 2010 at [60]; *Von Hannover v Germany*, 59320/00, Court, 24 June 2004 at [60].
[77] *Smith v. Collin* 439 US 916 (1978); see also the related case regarding the march itself: *National Socialist Party of America v. Village of Skokie* 432 US 43 (1977).

The town had enacted ordinances which outlawed speech that sought to intentionally incited hatred on the grounds of race, religion or nationality. Consequently, it refused permission when approached by the white supremacists. The ordinance was subsequently challenged by the neo Nazi group and held unconstitutional by the US District Court. An application for writ of certiorari to the Supreme Court was rejected. The Court held that no limitations on freedom of expression, especially political speech, on the grounds that the content was deeply offensive and hateful. The lack of engagement with the legal questions raised by the case was lamented by Justice Blackmun:

'I also feel that the present case affords the Court an opportunity to consider whether, in the context of the facts that this record appears to present, there is no limit whatsoever to the exercise of free speech. There indeed may be no such limit, but when citizens assert, not casually but with deep conviction, that the proposed demonstration is scheduled at a place and in a manner that is taunting and overwhelmingly offensive to the citizens of that place, that assertion, uncomfortable though it may be for judges, deserves to be examined.'[78]

This position was confirmed and expanded in the case of *RAV v. City of St. Paul*.[79] In the case, the Supreme Court overturned the conviction of a teenager for burning a cross on the lawn of an African American family under a hate crime ordinance. Writing for the majority, Justice Scalia held that:

'(…) the reason why fighting words are categorically excluded from the protection of the First Amendment is not that their content communicates any particular idea, but that their content embodies a particularly intolerable (and socially unnecessary) mode of expressing whatever idea the speaker wishes to convey. St. Paul has not singled out an especially offensive mode of expression—it has not, for example, selected for prohibition only those fighting words that communicate ideas in a threatening (as opposed to a merely obnoxious) manner. Rather, it has proscribed fighting words of whatever manner that communicate messages of racial, gender, or religious intolerance. Selectivity of this sort creates the possibility that the city is seeking to handicap the expression of particular ideas. That possibility would alone be enough to render the ordinance presumptively invalid, but St. Paul's comments and concessions in this case elevate the possibility to a certainty.'[80]

The ordinance was thus struck down for being content based, as it proscribed only certain expressions, but not others. This content neutral approach was confirmed by the Court in *Snyder v. Phelps*.[81] The case involved the picketing of the funeral of a US serviceman by members of the Westboro Baptist Church. Snyder's son, an active US serviceman, had been killed in Iraq following a vehicle accident. The religious group routinely picketed such funerals to protest what they perceived as the increasing tolerance of homosexuality in the US. Synder sued

[78] *Smith v. Collin* 439 US 916 (1978) at [919].
[79] *RAV v. City of St. Paul* 505 US 377 (1992).
[80] Ibid at [393].
[81] *Snyder v. Phelps* 562 US 443 (2011)

members of the Phelps family, who make up the vast majority of the Westboro congregation, for defamation and emotional distress. When the case reached the Supreme Court, it held that the statements of Westboro members were entitled to First Amendment protection and confirmed the appellate court decision to overturn a lower court ruling which had awarded Snyder compensation.

### *3. Incitement*

The case law of the Supreme Court on the extent to which expressions that seek to incite violence may be restricted and become the object of punishment for individuals has seen several stages of development and the application of a variety of tests and standards. It is not necessary for present purposes to offer a comprehensive account of the doctrinal history of this exception. It suffices to say at this point that the test currently employed by the Supreme Court is far from having a long pedigree in the case law. For much of its long history, the Supreme Court was comfortable and condoned limitations on free speech that were significantly more restrictive than the current doctrine. In that sense, the case law was long closer to the view taken by its European counterpart, and consequentially far from what might be seen as a certain free speech absolutism.

For years, the Supreme Court followed a balancing test that weighed the interests in free speech with the legitimate government interest to prevent harm. This led to decisions that upheld the convictions and other sanctions against pacifists and Communists,[82] but also overturned others concerning Jehovah's Witnesses and trade unionists.[83] As late as the early 1960s, the Court was openly applying a balancing test to free speech. The case of *Dennis v. United States* concerned convictions under a law that criminalised the advocacy of overthrow by force and violence of the US government.[84] The Supreme Court upheld the convictions finding that 'speech is not an absolute, above and beyond control by the legislature when its judgment, subject to review here, is that certain kinds of speech are so undesirable as to warrant criminal sanction.'[85] Crucially, it was not relevant for the purposes of upholding the convictions whether the overthrow was likely to succeed, the standard adopted was whether the gravity of the evil, regardless of its

[82] *Schenck v United States* 249 US 47 (1919) [distributing leaflets under provisions that criminalised the disrupting the recruitment of military personnel]; *Dennis v United States* 341 US 494 (1951) [advocating the overthrow by force and violence of the US government, regardless of improbability of success]; *Konigsberg v State Bar of California* 366 US 36 (1961) [refusal to answer question about Communist party affiliation at bar entrance exam].
[83] *Thornhill v Alabama* 310 US 88 (1940) [publicising the fact of an on-going labour dispute on premises of employer]; *Cantwell v Connecticut* 310 US 296 (1940) [publicly playing a record with inflammatory content].
[84] *Dennis v United States* 341 US 494 (1951).
[85] Ibid at [508].

improbability, justified the infringement of freedom of speech.[86] This was last confirmed by the Court in the case of *Konigsberg v. State Bar of California*. The Court held that

'(…) general regulatory statutes, not intended to control the content of speech but incidentally limiting its unfettered exercise, have not been regarded as the type of law the First or Fourteenth Amendment forbade Congress or the States to pass, when they have been found justified by subordinating valid governmental interests, a prerequisite to constitutionality which has necessarily involved a weighing of the governmental interest involved.'[87]

The case concerned the refusal of the California Bar to admit Konigsberg, as he had declined to answer any questions pertaining to membership in the Communist Party, asserting only that he did not believe in the violent overthrow of government and had never knowingly been a member of an organisation advocating violence.

The balancing test was effectively and abruptly abandoned in the late 1960s in a case overturning the criminal conviction of a white supremacist: *Brandenburg v Ohio*. Brandenburg, a KKK leader in rural Ohio was convicted for advocating violence following a speech given at a rally where he suggested that if the suppression of the white race through government were to continue, violent action might be necessary. The Supreme Court reversed the conviction, finding that the punishment of abstract calls for violence or unlawful action was incompatible with the First Amendment. The Court for the first time formulated a test which given the backdrop of previous cases might be considered a fairly absolutist position. It held that

'(…) constitutional guarantees of free speech and free press do not permit a State to forbid or proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action.'[88]

The Court thus established the now commonly applied three prong test to establishing the constitutionality of interference with free speech. Violence or lawless action must first be intended by the speaker, the expression must advocate for imminent violence or lawless action and such action must be likely.[89] While the *Brandenburg* test remains the starting point for expressions that advocate or in themselves constitute the commission of criminal acts, the Supreme Court has not had much occasion since to flesh out the contours of the test.[90]

---

[86] Ibid at [510].
[87] *Konigsberg v State Bar of California* 366 US 36 (1961) at [50] and [51].
[88] *Brandenburg v Ohio* 395 US 444 (1969).
[89] Confirmed in: *Hess v Indiana* 414 US 105 (1973).
[90] *Brandenburg v Ohio* 395 US 444 (1969).

### *4. Private limitations*

Individuals do not enjoy protections under the First Amendment against anything other than state action. Therefore, for instance the use of private property can be regulated at will by the owner, who may impose restrictions based on viewpoint, as well as on the time, place and manner of expression without having to justify the decision. In principle, this also applies to private property that is generally open to the public: the classic examples are shopping malls where applicants have in the past attempted to secure the right to protest and distribute leaflets.

This general position has not gone unchallenged and has led to the application of the First Amendment in some limited circumstances. In the case of *Marsh v. Alabama*, the Supreme Court struck down a trespassing statute invoked to prevent the distribution of religious material on a privately-owned company town sidewalk.[91] It held that the sidewalk had been dedicated to public use, reasoning that 'the more an owner, for his advantage, opens up his property for use by the public in general, the more do his rights become circumscribed by the statutory and constitutional rights of those who use it.'[92] This was expanded in *Food Employees v. Logan Valley Plaza*, where the Supreme Court found that the privately owned mall in question constituted 'the functional equivalent of the business district of Chickasaw involved in Marsh' and hence rejected the application of trespassing laws against peaceful picketing on the grounds of the First Amendment right to free speech.[93]

However, these decisions were significantly narrowed in subsequent cases. Paradoxically, this backlash was based to a significant extent on the robust protection offered under the First Amendment for refusing to exercise speech. It is generally unconstitutional for the government to require expression of a belief. The more prominent cases revolve around individuals refusing to pledge allegiance to the flag of the United States. In the case of *West Virginia State Board of Education v. Barnette* the Supreme Court held that a requirement to join a flag salute ceremony and pledge allegiance amounted to a compulsion of students to declare a belief.[94] In applying the at the time prevalent balancing test, the Court held that '(…) remaining passive during a flag salute ritual creates a clear and present danger that would justify an effort even to muffle expression',[95] a burden the state failed to meet. The Supreme Court thereby expressly overruled

[91] *Marsh v Alabama* 326 US 501 (1946).
[92] Ibid.
[93] *Food Employees v Logan Valley Plaza, Inc* 391 US 308 (1968).
[94] *West Virginia State Board of Education v Barnette* 319 US 624 (1943) at [633].
[95] Ibid at [634].

its previous ruling in *Minersville School District v. Gobitis*, where it had found no interference with the First Amendment under essentially the same factual circumstances.[96]

The *Barnette* case principle was applied to strengthen protections of private property owners against those who wish to make use of the property to exercise freedom of speech. It limited access to private property, initially requiring that the speech is related to the place where the expression is to take place, and ultimately holding that free speech guarantees do not apply to private property at all. In the case of *Lloyd Corporation, Ltd. v. Tanner* the Supreme Court held that where the expression in question is unrelated to the operations of the private place the protections of the First Amendment do not apply, especially when there are alternative fora available.[97] Tanner was distributing anti-war leaflets and had the opportunity to make use of the public sidewalk outside the mall for his activities, as this was owned by the local city. This was a key factor distinguishing this case from the previously mentioned *Food Employees* decision, where the expression was directly related to the shopping centre and no 'reasonable opportunities to convey their message (…) were available' to the protestors. The case firmly rejected equating public property with private property, even where it is intended for public use. This reasoning was confirmed in *Hudgens v. National Labor Relations Board*, where the Supreme Court held that the guarantees to free speech do not apply to private shopping centres.[98]

The Supreme Court further confirmed that this was effectively a debate over competing First Amendment rights, rather than one pitting the First Amendment against property rights. In *Wooley v. Maynard* a state law that required motorists to display licence plates with the slogan 'Live free or die' on their vehicles was struck down by a split court.[99] A unanimous court further found in *Pruneyard Shopping Ctr v. Robins* that First Amendment rights could not compel a private shopping mall to allow solicitation for signatures in a political campaign.[100] However, according to the Supreme Court, the US Constitution did not bar the Constitution of California from extending freedom of expression protections further. Under the California Constitution individuals were entitled to peacefully exercise their right to free speech in parts of private shopping centres regularly held open to the public. The Supreme Court thus established the principle that state constitutions may expand protections provided that the expansions do not

[96] *Minersville School District v Gobitis* 310 US 586 (1940) at [642].
[97] *Lloyd Corp v Tanner* 407 U S 551 (1972).
[98] *Hudgens v. National Labor Relations Board (No. 74-773)* 424 US 507 (1976); *Pruneyard Shopping Ctr v Robins* 447 US 74 (1980).
[99] *Wooley v Maynard* 430 US 705 (1977)
[100] *Pruneyard Shopping Ctr v Robins* 447 US 74 (1980).

themselves infringe on other protections. However, in doing so, it also rejected the argument that the extension of free speech protections amounted to an infringement of property rights of shopping mall owners. The Supreme Court thus confirmed that the issue was primarily one of the extent to which the First Amendment extended protections over the competing free speech rights of private property owners, and less a direct question of property rights as such. The distinction may appear nuanced and semantic at first blush, but it is an important conceptual difference in the way such constitutional complaints are structured.

## D. CONCLUSION

The discussion of the ECHR and US Supreme Court case law has demonstrated that the general approach to freedom of expression significantly diverges in both jurisdictions. In Europe, the ECHR recognises a need to balance state limitations on freedom of expression with competing interests and accepts that under certain circumstances the protections must be extended to interferences from private parties. This generally opens the door to laws such as NetzDG which primarily seeks to force social media platforms to adopt a complaints management system that can effectively enforce existing criminal law provisions. Nonetheless, both German constitutional law and the ECHR recognize the crucial importance of freedom of expression for a free and democratic society. Limitations are rightly treated with scepticism and are based on narrowly tailored legal requirements. Under both jurisdictions the core issue is whether the measures lead to overblocking of otherwise legal content. Should this occur, regulations are unlikely to be upheld. What little data is currently available does however not support this commonly expressed fear, but it is too early to draw firm conclusions. At the very least, NetzDG might serve as a starting point for regulatory frameworks under consideration in other European countries.

The same cannot be said for the US. The case law of the Supreme Court has gone through several distinct phases, not all of them were necessarily as absolutist on freedom of speech under the First Amendment. The current modern interpretation is however fairly unequivocal. Limitations are generally viewed with scepticism and any attempt to sanction the expression of a particular view are unlikely to survive Supreme Court scrutiny. Exceptions are only permissible under narrow circumstances, for instance direct incitement to violence, fighting words and obscenity. Overall, the constitutional position in the US makes NetzDG a less useful model for US regulation.

This scepticism with respect to state intervention contrasts with the disinterested approach to limitations imposed by private entities. There have been very few cases dealing with the scope of the First Amendment and the cases that emerged generally do not apply strong protections. This is crucial to the debate specifically surrounding social media, because such platforms are almost exclusively run by private entities. Paradoxically, it may be the European approach that in practice affords greater protection to freedom of expression online, as it does not dismiss out of hand the idea that such entities have certain obligations.

While Europe and the US may at first appear to strongly diverge, community standards, for instance those of Facebook, are largely compatible with limitations that European jurisdictions generally require, while equally not falling afoul of the First Amendment. Granted, the similar conclusions are based on markedly different perspectives on freedom of expression. Nonetheless, community standards developed through public consultation in both jurisdictions can form the basis for a fruitful and legally viable discussion on appropriate limitations for globally used platforms. Naturally, the legal endorsement of both jurisdictions is limited and does not come without caveats.

In the US the debate revolves more strongly around the issue of 'fake news' than hate speech, and any limitations are contingent on remaining (formally speaking) voluntary. This in turn highlights the gravest weakness of the prevalent US approach: freedom of expression on social media is almost entirely contingent on the cooperation and goodwill of private entities that are ultimately motivated by monetisation of content and profit, not the public good. It also explains the default approach from political activists with respect to these platforms: in the US much energy is focused on applying pressure to corporate entities to enshrine the desired stance on free speech in the terms and conditions of use.

By contrast, European jurisdictions tend to be concerned more with the lack of consistent and effective enforcement of already existing limitations on free speech. While the call for regulation in Europe generally finds a receptive legal framework, there are good reasons to be cautious: imposing further limitations and developing obligations on private entities carries tangible risks. Not all speech that offends and causes outrage can or indeed should be sanctioned. The danger is that states (and by extension private entities) restrict expression too severely, thereby choking off an important space for public debate. It would be overly optimistic to expect that stifling harmful expressions alone reduces the pervasiveness of the beliefs that inform them or an individual's willingness to act on them. The ability to voice harmful and

controversial, perhaps even bigoted views, can be an important first step in fully appreciating and ultimately changing harmful opinions.

In any case, legal frameworks and arguments can only influence the debate to a certain extent. Regardless of whether the First Amendment directly applies to private entities, the expectations of US users and the default approach of social media companies (who are chiefly US based) is undoubtedly informed First Amendment principles. Likewise, European attitudes to freedom of expression are informed by the more restrictive environments that already exist, regardless of whether it is wise to extend wholesale the duties placed on states to private entities. Social media platforms have brought about undeniable benefits and while the current mood (rightly) seeks to correct years of regulatory *laissez faire*, we should strive not to overtreat the malaise we have diagnosed.