Competent Third Parties and Content Moderation on Platforms: Potentials of Independent Decision-Making Bodies From A Governance Structure Perspective

Author(s): Amélie Heldt and Stephan Dreyer

Source: *Journal of Information Policy*, 2021, Vol. 11 (2021), pp. 266-300

Published by: Penn State University Press

Stable URL: https://www.jstor.org/stable/10.5325/jinfopoli.11.2021.0266

COMPETENT THIRD PARTIES AND CONTENT MODERATION ON PLATFORMS

Potentials of Independent Decision-Making Bodies From A Governance Structure Perspective

Amélie Heldt and Stephan Dreyer

ABSTRACT

After many years of much-criticized opacity in the field of content moderation, social media platforms are now opening up to a dialogue with users and policymakers. Until now, liability frameworks in the United States and in the European Union (EU) have set incentives for platforms not to monitor user-generated content—an increasingly contested model that has led to (inter alia) practices and policies of noncontainment. Following discussions on platform power over online speech and how contentious content benefits the attention economy, there is an observable shift toward stricter content moderation duties in addition to more responsibility with regard to content. Nevertheless, much remains unsolved: the legitimacy of platforms' content moderation rules and decisions is still questioned. The platforms' power over the vast majority of communication in the digital sphere is still difficult to grasp because of its nature as private, yet often perceived as public. To address this issue, we use a governance structure perspective to identify potential regulatory advantages of establishing cross-platform external bodies for content moderation, ultimately aiming at providing insights about the opportunities and limitations of such a model.

Keywords: content moderation, social media platforms, governance, intermediary liability

The current debate around social media platforms is not only on content moderation, it involves a broader picture including, inter alia, the data

Amélie Heldt: Leibniz-Institute for Media Research, Hans-Bredow-Institut, Hamburg, Germany
Stephan Dreyer: Leibniz-Institute for Media Research, Hans-Bredow-Institut, Hamburg, Germany
DOI: https://doi.org/10.5325/jinfopoli.11.2021.0266

on which the attention economy is built, the differing motifs of ad-based platforms and users driven by their individual interests, the potential influence that engaging content can have on users, and how algorithms handle content that is not only engaging but moreover contentious and toxic. It is a challenge in itself to keep track and to see the whole picture while being aware of the specificities of each of these issues. This article responds to that need by using the analytical approach of focusing on the underlying governance structures, giving an overview of the challenges and obstacles encountered with regards to content moderation by states and platforms so far, and offering guidelines for a proper use of new institutions oriented toward general interests in this context. In our contribution, we analyze the guiding opportunities arising from a competent third-party system involving technology in the context of content moderation from a governance structure perspective, showing out the advantages of such a scenario compared to the status quo, and pointing out the requirements. Ultimately, we conclude that, under the current circumstances, a competent third party independent from both governmental and corporate interests that is equipped with powers to decide on content moderation matters could be the way forward, provided that the challenging requirements are met.

Background: Liability as a Janus Face in Platform Governance

The principle of platforms' limited legal liability was initially based on the idea that they perform a service of distribution, not of publishing or performing. This general idea was spelled out in a US Supreme Court landmark case (*Smith v. California*) and later translated into national legislation, for example, Section 230 of the Communications Act of 1934 (47 U.S.C. 230), commonly known as the Communications Decency Act in the United States. Similar approaches can be found in Article 14 of the European Union (EU) e-Commerce-Directive or in Section 79 of the Indian Information Technology Act. The generic term for social media platforms, search engines, and microblogging services—information intermediaries—reflects the starting point of the discussion around platform governance, that is, intermediaries function as a conduit or a pipe. They do not edit content or produce their own direct speech. The initial idea was also to give these new businesses an opportunity to grow without being fully affected by the burden of far-reaching monitoring obligations. At the beginning of the "social web" also known as "web 2.0," the opportunities

of new communication networks raised high expectations.[1] Intermediaries were perceived as a medium of connecting users, allowing each of them to share content both within their networks and publicly, and enhancing global information exchanges. Unlike traditional media outlets, intermediaries are—at least in principle—not obliged to investigate content before allowing it to be uploaded. They may moderate content and are supposed to do so in specific cases but are in general exempted from any duty to proactively looking out for unlawful user-generated content (hereinafter UGC). According to EU law, intermediaries are only liable for infringing content of which they have knowledge. If a platform provider systematically investigates UGC, it might be held responsible for all content due to these monitoring measures. The legal framework set out by the e-Commerce-Directive produces a systemic incentive not to establish any form of automated content monitoring; Article 15 explicitly prohibits the Member States the establishment of rules obliging platform providers to do so. This is surprising, as usually liability is used as a resource to control active measures of private parties.

This type of (governmental) regulatory framework heavily molded the image of unaccountable social media platforms that rather leave problematic UGC online than take it down. In practice, it depends on the platforms: how much they moderate content, according to which standards or rules, and by which means.[2] Inevitably, larger platforms have to deal with a bigger variety of users and of UGC, which in turn pushes them to adopt a distinct approach toward content moderation.[3] When confronted with problematic UGC such as defamation, libel, extremism, propaganda, and other forms of disinformation, child sexual abuse material, but also copyright infringements, once informed platforms have to act and cannot rely on the provisions mentioned earlier. They will either stick to their relative immunity as provided by limited liability laws (e.g., Reddit, 4chan), or engage in more active forms of content moderation (Facebook, YouTube, Twitter, etc.). The interests at stake with content moderation and with the governance of content moderation are third-party rights such as personality rights, copyrights, and, at a higher level, public safety and order, or even democratic values including autonomous formation of political opinions,

1. Tucker et al.

2. Gorwa, Binns, and Katzenbach.

3. In this article, "content moderation" generally means content control on social media platforms, performed by nonstate actors unless otherwise noted.

free public discourse, and election integrity.[4] For almost two decades, platforms hosting UGC were only subject to this broad regulatory framework, according to which they could moderate content according to their own rules.[5] Private ordering, that is, the rulemaking and enforcement thereof by private actors, is primarily motivated by reduced costs for the state and enhanced flexibility and efficiency for private actors.[6] Lawmakers in the United States and the EU adopted this model and allowed platforms to govern UGC by means of their respective terms and conditions.[7] In light of the spread of contentious UGC over social networks and the potential and observable harms this content does to the individual and societal interests in general, the platforms' relative immunity is increasingly contested and their relevance for public communication is more and more highlighted in regulatory debates.[8]

This criticism is mostly based on the findings that have shown over the last years that some platforms are not as content-neutral as they claim to be.[9] Not only does their business model rely on users to generate content, moreover the algorithmic prioritization system prefers content with high user engagement. That strategy comes at a high price when the most engaging content is also the most contentious.[10] Plus, this puts their role as mere distributors into question, as they select, filter, and prioritize content according to the alleged interests of their users. In the same time, platforms that aim at bringing together larger parts of society all over the world and not limiting their target audience (e.g., to certain topics such as themed forums) play an active role in moderating UGC—even if their internal policies and decision-making processes in the area have only been revealed to the public in recent years.[11] It turns out that the "opacity of private

4. Collins, Marichal, and Neve; Tambini; Batorski and Grzywińska.

5. Schwarcz, 320.

6. Ibid., 327–28.

7. Belli and Venturini.

8. See, for example, the public announcements of the US Department of Justice to discuss Sec. 230 CDA ("and whether improvements of the law should be made"), 30 January 2020, via https://www.justice.gov/opa/pr/department-justice-hold-workshop-section-230-communications-decency-act and of the EU Commission regarding the EU Digital Services Act: Ursula v. d. Leyen, "A Union that strives for more—My agenda for Europe" (2019), p. 13, via https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf; De Gregorio, "Democratising Online Content," 105374.

9. Gillespie, "The Politics of 'Platforms'"; Gillespie, "The Platform Metaphor."

10. Crockett; Deibert; Tufekci.

11. Cf. Kettemann and Schulz.

speech practices"[12] is an important concern because the lack of transparency prevents all checks and balances that we apply when traditionally examining state action that affects speech. Indeed, regulation is not per se a legislative act: any intervention that "links ordering processes with explicit objectives and measures" may be considered regulation.[13] In a narrow sense, regulation or regulatory frameworks that are "issued for the purpose of controlling the manner in which private and public enterprises conduct their operations"[14] are usually associated with legislative or state authorities' interventions,[15] as distinguished from forms of self-regulation and private ordering.[16] Co-regulatory frameworks usually combine regulatory frameworks and state oversight with such types of self-regulation or private ordering.[17] Research has shown that social media platforms acting on a global market have their very own set of standards and rules and enforce them—sometimes beyond what states would be allowed to do.[18] With regard to the principle of contractual freedom this is not a problem per se, but it raises the central question regarding the role and functions of intermediaries for individuals practicing their fundamental rights to communicate and inform on one hand, and their supra-individual function for public communication in democratic societies on the other hand. This observation goes hand in hand with legislators wanting intermediaries to cope with the growing responsibilities that come with their power over digital communication.[19] What can be observed in the policy discourse in the last years is a shift (back) from a *laissez-faire* approach with unobtrusive legal provisions and rather liberal governance approaches to forms of stronger, more legislation-focused, and more demanding media regulation.

Legislators in the EU have endorsed stricter policies when it comes to online intermediaries than in the beginning of the 2000s when drafting Art. 14 e-Commerce Directive. Search engines have to delete personal information (Art. 17 EU General Data Protection Regulation (GDPR):

12. Citron.

13. Hofmann, Katzenbach, and Gollatz.

14. Majone.

15. Baron ("Regulation involves government intervention in markets in response to some combination of normative objectives and private interests reflected through politics."); cf. Puppis.

16. Schulz and Held.

17. Ibid.

18. Gillespie, Custodians of the Internet: Platforms; Klonick, "The New Governors"; Suzor; Roberts, Behind the Screen: Content Moderation.

19. Kinstler; Echikson and Knodt.

right to be forgotten), social media platforms are expected to remove unlawful content within short time spans, for example, in cases of child sexual abuse material, terrorist propaganda, or copyright infringements. Under Section 3 NetzDG (German Network Enforcement Act),[20] obviously unlawful content has to be removed within 24 hours. Under Article 4 of the proposed EU Regulation on preventing the dissemination of terrorist content,[21] an online "hosting service provider" has to remove "terrorist content" within one hour. Platforms are expected to react almost immediately when it comes to unwanted content, not least to avoid its spreading throughout the vaster Internet (beyond the networks of a single platform). Moreover, the EU is about to draft and adopt a new set of rules that will target the issues mentioned here and update the E-Commerce-Directive, that is, the Digital Services Act (DSA). It is expected to bring more procedural obligations for content moderation, more transparency for content curation, and a more aligned approach at the European level. According to the documents published by the relevant committees of the EU Parliament, the DSA will have profound implications for the platforms but will also harmonize the legal regime across Europe. Actually, the LIBE (Committee on Civil Liberties, Justice and Home Affairs) committee's report proposes an independent body to exercise effective oversight—an idea very much in line with this article's core argument.[22] In addition, the EU Commission announced a European Democracy Action Plan, supposed to "ensure that citizens are able to participate in the democratic system through informed decision-making free from unlawful interference and manipulation."[23] It is important to note that the purposes of regulatory interventions have taken an observable shift from ensuring individual rights to discussing countermeasures against risks for both individual rights and supra-individual aims like societal values or public interest.[24]

These legislative activities show that lawmakers believe platforms should protect the EU's core values. However, putting the platforms in charge of a faster and more effective removal of unlawful or unwanted content is not as simple as it appears when reading the laws mentioned earlier, with

20. As well as under Section 1 Loi Avia (French Law against Online Hate Speech), but the law was annulled.

21. European Commission.

22. Peeters.

23. EU Commission.

24. Napoli.

content moderators being responsible for interpreting vague legal terms.[25] The enforcement of laws that restrict freedom of speech via private actors is a source of concern.[26] It comes along with a shift of power toward nonstate actors in matters that are *a priori* in the domain of the judiciary. Moreover, such laws can be a backdoor to collateral censorship when states impose the removal of speech in a way that they themselves would not be allowed to do. The fear of collateral regulation or censorship by proxy is not new[27] but it might be intensified when the state hands over significant duties such as judicial review to nonstate actors and, additionally, is not in full control of the way the enforcement is actually performed in terms of legal interpretation. Sanctioning platforms when they do not comply with removal obligations within a short time span also bears the risk of incentivizing removal of borderline content although it is, in fact, not unlawful. On the one hand, pressuring platforms to react when users complain about harmful content via financial sanctions is effective, but on the other, it might come along with consequences that go against the initial rationale of the regulatory intent.

All in all, content moderation on social media platforms is a manysided topic that cannot be summarized into one problem,[28] hence it cannot be answered by a one-size-fits-all solution. Neither governments nor the companies running social media platforms have found an appropriate agenda covering all aspects. Given that social media platforms operate on a global scale and that their users are spread all over the world, hence subject to different jurisdictions, there is so far neither a single regulatory framework that would match, nor is there supranational agreement regarding a binding legal international minimal standard when it comes to recognizing the human rights involved.[29] The current legal framework generally stipulates that a platform provider's knowledge of any infringing content results in liability, thus setting up an incentive structure where systematic content monitoring leads to stricter responsibility. This is the underlying dilemma of today's platform liability framework. Moderating content from different cultures and regions are tasks that come at high costs (economic and psychological) when performed by human moderators.[30] To address this

25. Ruckenstein and Turunen.
26. Kaye.
27. Birnhack and Elkin-Koren; Kreimer.
28. Grimmelmann; Stalla-Bourdillon and Thorburn.
29. Council of Europe.
30. Roberts, "Commercial Content Moderation."

issue, we need to take a closer look at the chances (and risks) offered by an independent actor, including a look at how such changes in the governance structure might support advantages with regard to cross-platform content moderation and technology-driven forms of platform governance. Although technology might support and/or perform external content moderation, paying off in terms of scale and efficiency, it would need to be designed in a very cautious way to not reproduce current shortcomings in this area, potentially infringing the human rights and principles mentioned earlier.[31]

Online Speech and Power Shifts in Public Communication

In the digital sphere, protecting democratic principles and freedom of speech as well as access to information, media freedoms, and so on, is no longer a matter of protecting solely from the coercive power of the state.[32] It has to incorporate the decentralization of power and, therefore, power shifts toward new private actors. Thus, the environment for public communication on the Internet has similarities with the analogue world, but is different when it comes to speech because it more often involves three players[33]; at least it is the case when we focus on online intermediaries. There is the state, the user (who is at the same time a communicator and a recipient), and lastly, the platform providers. The latter are private actors that have in some cases acquired a substantive power over communication for individuals and society as a whole, positioning themselves closer to governments with regards to their normative power.[34] Governance approaches were drafted and their potential implementation analyzed to make sense of the growing importance of regulation through private actors. These shifts of power started before the rise of social media platforms, that is, in the course of globalization and the emergence of vertical and horizontal self-regulatory arrangements between state and nonstate actors.[35] Black describes the decentralization of regulation as follows: "So

31. Previous analyses of this problem include: Citron; Perel (Filmar) and Elkin-Koren, "Separation of Functions for AI."

32. i.a. Yemini.

33. Balkin, "Free Speech Is a Triangle."

34. Tushnet; Citron and Quinta; Helberger, Kleinen-von Königslöw, and van der Noll; Katzenbach et al.

35. Abbott and Snidal, "The Governance Triangle: Regulatory Standards"; Black, 108.

complexity, fragmentation and construction of knowledge, fragmentation of the exercise of power and control, autonomy, interactions and interdependencies, and the collapse of the public/private distinction: all are elements of the composite "decentred understanding" of regulation."[36] The description of Black has to a large extent become reality with the Internet, and the growth of online intermediaries' services has reinforced this pattern. It has in parts replaced hard law instruments where self-regulation might have appeared as a white hope in many aspects.

This position of intermediaries as the new gatekeepers of the information society begs the question of their role when it comes to their users' rights. Legislators and authorities decide what is legal and what is not, but they increasingly delegate power to or rely on platforms to enforce these laws online while leaving room for the platforms' own sets of standards and rules. The latter are more convenient to enforce because it is at each platforms' discretion whether or not to ban certain types of content, and whether to remove it locally or globally. Lessig's "code is law" still stands for the normative power by means of platform architecture.[37] Although this formula is still valid, lawmakers no longer leave it up to platforms' discretion to develop rules within a self-regulatory framework. The lawmakers' trust placed in self-regulation has diminished over the past years[38] and, at least in the EU, they are increasingly drafting regulatory frameworks that stipulate "clear" rules and "hard" consequences.[39] In their overview of the different governance strategies, Latzer et al. point at the weak points of self-regulation.[40] Indeed, the governance of online intermediaries has been criticized for many years for not being sufficiently effective and coherent.[41] Gasser and Schulz identified three types of situations in their international study of online intermediaries' governance: countries that lack specific intermediary governance frameworks, countries with existing and differentiated specific frameworks, and countries with emerging frameworks.[42] The identified deficiencies were mainly connected to the heterogeneity of intermediaries' regulation and the lack of awareness of the possibly negative

36. Black, 117.

37. Lessig, 1.

38. Cammaerts and Mansell, 142–43.

39. De Gregorio; see also: reports from the JURI, LIBE and IMCO Committees of the EU Parliament for a Digital Services Act.

40. Latzer, Just, and Saurwein.

41. Gasser and Schulz, 27.

42. Ibid., 16.

consequences on users' fundamental rights.[43] All in all, scholars agree on the importance of adopting a global perspective while still bearing in mind the regional context when designing instruments of content moderation that also take human rights into account.[44]

For today's major platform companies, the governance landscape in the EU and the United States is quite different with regards to the lines they are taking.[45] The EU's Member States tackle the problem of harmful content by making platforms cope with local laws and reducing their immunity. Germany set off the regulatory wave with the Network Enforcement Act (NetzDG) in 2017, a law under which platforms are obliged to ensure user-friendly complaint mechanisms and the quick removal of unlawful content. France wants to ameliorate the enforcement of its laws against information manipulation in times of political campaigning by installing a new preliminary ruling procedure and transparency obligations for sponsored content.[46] On the supranational level, the EU adopted several codes of conduct tackling online hate speech and extremist content. However, these regulations might not solve the issue of moderating content in a proportionate manner, that is, to limit the speech restraints to the smallest possible while protecting other rights as well. As Grimmelman puts it: "No community is ever perfectly open or perfectly closed; moderation always takes place somewhere in between."[47]

Laying out the Governance Structure: Spheres of Efficacy

Traditional legal perspectives take a normative focus when analyzing a specific policy field, focusing on norms and provisions manifested in legal texts and jurisprudence. The consequences these norms have or might have on the addressees as well as on other relevant actors within the field of application are, however, a traditional topic for the sociology of law. Hart's approach to a "Concept of Law" was the first one aiming at bringing together these perspectives: Using a stance of a "descriptive sociology of law," he developed the social-scientific concept of a governance structure that looks at socio-legal compounds with perspectives on the behavior of

43. Ibid., 17.
44. Ibid.; Latzer, Just, and Saurwein, 390; Gorwa, 15.
45. Gorwa, 6.
46. Heldt, "Von der Schwierigkeit, Fake News."
47. Grimmelmann, 109.

the players within a specific area and its entanglement with normative, specifically legal contexts.[48] The concept has been further elaborated by Kornhauser to be able to close the gap "between the institutional focus of governance structures and the normative focus of the jurisprudential debate," taking into account formal and informal incentives for institutions and their behavior.[49] The two important features within the concept of governance structures are the degree of institutional differentiation and the motivating factors underlying the institutional design, especially in cases where a specific institutional figuration needs to resolve the problems of recognition, adaptation, and enforcement where primary (legal) rules are either questioned, where conditions change rapidly, or societies become less homogeneous.[50] In the last two decades, legal scientists have found the concept of governance structures as a fruitful analytical approach in times of social differentiation and challenged traditional regulatory concepts.[51] In complex legislative environments consisting of hard law, soft law, informal means, and self- or co-regulatory initiatives, the largely distinct structures of legislation, monitoring, and enforcement bear different, often complex, relationships with the addressees of regulatory measures, not only, but also depending on their respective organizational structures and differing motivational processes.

Before this background, we use the concept of governance structures as a theoretical starting point for analyzing the actors in the three vertices of the triangle, their formal and informal incentives for specific behavior, and, subsequently, the degree of institutional differentiation and dependencies between them when it comes to the issue of moderating UGC. By applying the perspective of governance structures to the described triangle, we are able to describe the institutional figurations between the three players and their potential influence on resolving the challenges of content moderation (spheres of efficacy). Using the concept of governance structures, we can systematically explain the current stalemate situation in platform governance.

Formally, the state is the only legal authority in the online speech triangle we describe and, as such, has the monopoly on legitimate use of force, for example, by enforcing laws concerning illegal content. It shapes the

48. Hart.
49. Kornhauser, 375.
50. Hart, 91.
51. Dreyer, 120; Trute, Kühlers, and Pilniok.

legal framework for both the platform and the user in terms of illegality of content as well as of liability for illegal content, taking into account the knowledge and the degree of awareness of illegal content. The legislator's sphere of efficacy is broad and comprehensive—at the same time, it is guided and limited by individual rights, civil liberties, and constitutional principles, which might have a wide scope and, accordingly, limit the legislator in establishing and the authorities in enforcing existing legal provisions. When it comes to individual UGC on platforms the state runs into legal, technical, and factual access limitations. When content generated by users is uploaded on platforms using their infrastructure and services, such actions take place in the realm of an agreement between two private actors. The state may regulate the contexts and power relations between the private actors, but it is not allowed to specifically intervene between the two or to request direct access to nonpublic platform communication. When states chose to regulate speech (and when speech regulation is not forbidden by constitution), they are usually bound to the principle of common standards. Legal liability frameworks are usually being used as a resource of control here. When users communicate anonymously or are difficult to get hold of, the state will usually fall back to address platforms as subsequent access points for all forms of governance.[52] As seen earlier, the state will usually start to demand more "social responsibility" from the platforms, later on putting more pressure to self-regulate on the platforms by threatening them with formal legislative provisions. In cases where self-regulatory measures do not yield expected results, states may start to pass bills that directly regulate platforms with regard to their content moderation. This approach, however, comes with the risk of capturing platforms via expression-related legislation, and using them to enforce speech regulation in abusive ways in worst case scenarios. All in all, however, the state usually exercises rather indirect forms of control when it comes to content moderation on social media platforms. As stated earlier, legislators usually intervene with the purpose of safeguarding individuals' basic rights. A rather new issue is whether and how states should react to supra-individual or even societal risks emerging from platform-based content moderation. As communication on platforms and content moderation on scale can also have detrimental effects on public communication, social cohesion, or democratic opinion-formation, the question is whether the legislative handling of such risks also should be considered in this approach. As the

52. Elkin-Koren and Haber, 106.

governance structure approach focuses on the relevant actors and their spheres of efficacy, two aspects emerge regarding such societal risk mitigation: First, the powers of the State remain unchanged in this area; it will be thrown back to indirect forms of governing here, too (see above). The reason for this currently is that legal doctrine still has to provide a concept of positive legal obligations when it comes to legitimizing the State to safeguarding societal values. It remains to be seen whether such a doctrine would change the power structure between the State and the platforms. Second, there is not enough empirical data (yet) on the relevance of platforms for public communication and respectively alleged risks.[53] For now, we assume that extending the regulatory purposes from individual rights protection to ensuring societal values does not change the spheres of efficacy described here.

Platforms, on the other hand, are private companies that stand in a contractual or contractual-like relationship with users according to the principle of contractual freedoms (limited, for example, by data protection laws, copyright laws, consumer protection laws). Hence, they have direct access to all users' uploads, they can perform content analysis like file recognition, object detection, hash value-based detections, audio fingerprinting and analysis, action recognition, and alike.[54] They perform user tracking and behavior-based user profiling and conduct predictive analyses regarding the alleged interests and needs of their users.[55] With regard to unwanted content, platforms can, therefore, exclusively and finally decide to systematically identify and take-down content and to prevent reupload of identical or nearly identical material.[56] All in all, platforms exercise very direct forms of control on their users' speech on the basis of their community guidelines, hence, an extensive sphere of efficacy. When necessary, their rules are enforced by taking down content, algorithmically downgrading users' visibility, and suspending user accounts.[57]

The relationship between platforms and users is mostly unbalanced: platforms can de facto enforce any rule and, additionally, influence the users' communicative behavior and the exercise of their rights. This has led to a discourse about taking into consideration the complexity of this

53. Keller and Leerssen, 46; this is easier from an economic perspective, as Fagan (2018) has shown, Fagan, 393.

54. Cf. Gorwa, Binns, and Katzenbach.

55. Dijck and Poell, 10.

56. Heldt, "Upload-Filters: Bypassing Classical Concepts," para. 13 f.

57. Klonick, "The New Governors," 1630 f. How Are Platforms Governing?

relationship when lawmakers plan regulatory measures that could be detrimental to the platforms' business model. Because platforms aim at avoiding hard and external regulation at any cost they will self-regulate in lockstep. This combination can lead to a form of "unholy alliance" between the state and the platforms and have serious—sometimes even unnoticed by politics—consequences by collusion for the users' freedom of expression.

Last but not least, users. They are often perceived as the weakest player in the triangle, although they actually are the ones directly publishing, interacting with, and sometimes deleting their own content. They are the user base that a platform needs to function. In relation to both the state and the platform, users are the ones being governed. At the same time, platforms can seize the user base for identifying relevant content. Users can report problematic content by tagging or flagging, thus making the platform aware of it.[58] Some platforms have a moderation system where selected users or moderators are attributed authority when it comes to moderating content, for example, YouTube's or Reddit's "trusted flagger" programs. In a sense, such mechanisms are a crowd-sourced form of control (voting, tagging, reporting, or blocking),[59] and, thereby, the users can force a platform to react where it perhaps need not before.[60] Nonetheless, the sphere of efficacy of users is much more restricted than the legislators' and platforms'—due to their lack of control over the enactment and enforcement of rules, users are significantly more dependent on platforms and legislators than vice versa.

As the relation- and efficacy-related analytical dimensions have shown, the platforms are the ones with the most direct access to content at scale, including semi- or fully automated forms of content moderation. Generally, moderation *ex post* is perceived as the best solution when it comes to content moderation: it allows both fast publication of UGC and content control.[61] Nevertheless, taking down harmful content after its publication might still come too late with regards to the attention peak of propagated content. The popularity and "viral spread" of UGC can be due to content-based and other "content-agnostic" factors.[62] In cases of extremist or terror propaganda UGC, the major impact can happen within the first moments or days after publication, due to a so-called first mover

58. Crawford and Gillespie; Buni and Chemaly; Myers West, 4374.
59. Dreyer and Ziebarth, 534.
60. See supra. Part A, on the limits of platform liability.
61. Greis et al., 1435–38.
62. Borghol et al., 1.

advantage.[63] Also, the first numbers of views are perceived as a strong indicator of the popularity of online content.[64] The argument could be made that contentious content can be recognized by its "virality" and that an extended timeline for ex post moderation, that is, a procedure of notice and take-down would be too long for unwanted content, since its detrimental effects on both individual rights and societal interests might take place in the meantime. Additionally, the majority of users does not wish to be exposed to hate speech or toxic content—which could be an argument in favor of a faster removal in accordance with the law.[65]

Users expect platforms to act and to remove extremist content, as seen with the Christchurch livestream.[66] Nonetheless, one must be aware that too high incentives for platforms to delete content can be a threat for freedom of expression and information.[67] The effect of over-removal might be exponentiated as much as the immunity of platforms tends to be less comprehensive. More liability, less attractive to advertisers and users—there is a strong motivation to remove contentious UGC rather than keeping it published. Along those lines experts also red-flag laws stipulating an obligation for platforms to implement "pro-active measures" to detect unlawful content, since this could potentially lead to an unprecedented peak of private control over online speech.[68] Using automatic upload filters or other technological tools that analyze all content before its publication and without any grounds for suspicion is very likely to be considered unconstitutional if it is based on a (lacking) type of legislative framework.[69] Some scholars call for a broader interpretation of state action, that is, a stricter scrutiny for instruments of soft law when their implementation by private actors will factually lead to an impermissible restriction of fundamental rights.[70]

In sum, there is a constant back and forth between private ordering by social media platforms on the one hand, and regulatory frameworks that limit intermediary liability or force them to enforce laws on the other.

63. Ibid., 6.

64. Szabo and Huberman, 9.

65. See also Klonick, "The New Governors," 1625 Why Moderate At All?

66. Greis et al., 4; Gillespie, Custodians of the Internet, 8–9; Klonick, "The New Governors," 1625–30.

67. Kaye, 7; Citron, 1039, 1055.

68. Balkin, "Old School/New School," 2296–342; Kaye; Keller.

69. Heldt, "Upload-Filters: Bypassing Classical Concepts of Censorship?"

70. Birnhack and Elkin-Koren, 49 f.; Elkin-Koren and Haber, 107.

Both sides can be more or less restrictive, that is, for platforms: having a rather broad definition of unwanted content or as little intrusive content moderation policy as possible; or for state actors: strict constitutional proviso that only allow very narrow exceptions in the scope of protection of free speech or censorship friendly regulations usually recognizable by the use of vague legal terms. It is therefore not possible to assert which of private ordering or state regulation is the most effective and least harmful at the same time when it comes to the users' individual freedom. Hence, we have reached a stalemate situation right now when it comes to platform governance. To progress from here one needs to think beyond traditional forms of governance where a task is traditionally fulfilled by a single actor, entailing the risk of overlooking an important component. If one looks only at the individual actors, not enough attention is paid to the links they have between each other. This is where a governance structure perspective can show ways out of the current dilemma. Instead, it might be worth to look at collaborative forms of governance, including new and independent actors. This approach might facilitate identifying a potential governance structure in which the individual is not either user or citizen but both because the division between the contractual relationship between users and platforms on the one hand, and the societal position of individuals as human rights holders on the other, would be made void. Furthermore, the platforms would not become the mighty new regulatory bodies for online speech.

Concepts of Collaborative Forms of Governance

Internet regulation and platform governance are not the only fields where new answers have to be found for the relationships between state and non-state actors, and for the question whether there is a need for legislators to become active.[71] With regard to the latter, previous research was conducted in other fields of legislation. In environmental law, for example, there is less legislative activity to be observed when private actors commit to voluntary and binding rules, although private self-imposed rules do not automatically lead to a better protection of the environment. It is due to legislators tending to settle for less strict rules when these voluntary rules

71. For example, Loo, 1319 on the oversight of the NY Stock Exchange.

come from the private sector.[72] Private regulation can support aspirations of governmental policies but not replace them because they aim at different values than those protected by state actors in representative Western democracies.[73] Although private actors tend to prevent (new) state regulation, since it comes with (new) costs, the state is directly bound to both recognize human rights and—in some regions—safeguard these rights against limiting activities by third parties.[74] It is also an argument in favor of choosing constitutional principles such as the separation of powers to guide deliberations on future governance.[75]

This general observation does however not stand in the way of more collaborative forms of governance, which can be preferable to others as they bring together competencies from different actors.[76] Studies have shown that co-regulation, for instance, shows regulatory advantages in highly dynamic areas where expertise by the addressees is needed,[77] or that self-regulation can provide solutions to legislative stalemates due to highly politicized issues, even though the results of self-regulation have not been the most fruitful in the area of social media governance so far.[78] The field of content sharing and social networks is particularly starving for innovative governance concepts because of the shown triangle formed by the actors involved.

Collaborative governance might suggest a joint form of action but it should not be confused with multistakeholderism as it is not (only) about bringing together all actors involved in the matter of content moderation. Either way, the term "multistakeholderism" has been used as a proxy for alternative governance models that would involve more than one or two actors[79] and is therefore not so far away from what this article is looking for. Even if multistakeholderism has been discussed in relation to the platforms' nature, the question we want to address here is more of administrative nature, rather than a matter of intergovernmental organization.

In the range of collaborative models, one option can be to reinforce the collaboration between single actors. Besides the general issues observed in

72. Malhotra, Monin, and Tomz, 34.
73. Cammaerts and Mansell, 139.
74. Ibid.
75. Weber and Gunnarson, 37.
76. Abbott and Snidal, "The Governance Triangle," 28; Helberger, Pierson, and Poell, 1–14.
77. Schulz and Held, "Study on Co-Regulation Measures."
78. Cannataci and Mifsud Bonnici, 57; Hirsch, 439–64.
79. Raymond and DeNardis, 611.

cooperation-based governance structures, for example, power hegemonies reflecting in discourse outcome or forms of capturing,[80] the various combinations come with advantages and flaws, typically because they involve two of the three parties and destabilize the "triangle constellation" for better or worse, depending on the actor asked. That being said, it is probably not appropriate to call the current state of power relations within the triangle of governments, users and platforms "balanced" (see supra). Either way, none of the contemplated combinations has met the identified requirements of balanced content moderation so far, that is, a balanced approach of contractual freedom and constitutional principles. Formulated in an overexaggerated way, the status quo of the triangle has, so far, always led to an odd man out and it generally tends to be the user—and with him or her, her or his human rights. If we want to design new governance forms, we need a clear understanding of the pros and cons mentioned.[81]

When states and platforms work together it comes hand in hand with the risk of the state acting *via* these private actors, that is, collateral censorship, serious threats for individual rights, due process, and other important principles. A too close relationship between states and platforms is rightly perceived as too opaque and powerful.[82] Moving away from the state's responsibility, other problems might emerge. If platforms have the freedom of creating their own rules based on their contractual relationship with users, they might be driven by ad revenues ("attention economy") and only incidentally by the ideals of free expression and information.[83] Without formal governmental intervention, the platforms' room for interpretation of how to govern the digital sphere and both public and private communication within is extensively broad and left at their discretion.[84] Here, one can deplore that individuals act more as mere customers than active citizens although they could potentially influence not only the market but also their governments.[85] Sometimes, users are solicited and contribute, for example, by flagging unwanted content (YouTube), by moderating content in a bottom-up approach (Reddit), by installing and maintaining alternative dispute resolution models (Wikipedia), or by voting trustworthiness

80. Cammaerts and Mansell, 142–45.

81. The following sentences are summaries of contemplated scenarios and serve the purpose of briefly mentioning their respective downsides.

82. Birnhack and Elkin-Koren; Citron.

83. Marantz.

84. Langvardt.

85. Helberger, Pierson, and Poell, 4, 12.

of news (discarded Facebook idea).[86] When platforms and users initially cooperate on the subject of content moderation, the instruments used can be similar to those used in representative democracies. Rules and the way of enforcing them can be commonly agreed upon. However, it seems that the growth of user numbers comes hand in hand with less collaboration between platforms and users.[87] With the rise of user numbers, a system of peer-based moderation might no longer be an option for large platforms because it would lose its efficiency[88] or might be played by users acting collusively to misuse user-based reporting and moderation tools. Hence, the platform's size can coincide with the implementation of commercial content moderation.[89] The latter comes along with the additional challenge for commercial content moderators to tackle hate speech or other issues in other, less familiar cultures and contexts.[90]

Independent Third Parties as a Solution? Advantages from a Governance Perspective

As now elucidated, none of the solutions in place nowadays is compliant with constitutional principles nor viable in the long run as they respectively include a power disequilibrium.[91] If liabilities law and the intrinsic motivations of platforms tend to tilt toward paradox patterns of both too much and too little content moderation, the current governance structure does not seem able to cope with balancing rights and duties within the triangle.[92] That is why this article looks at the possibility of adding an additional player in the equation.[93] We are building this idea on the hypothesis that by establishing independent bodies that take content-related decisions instead of platforms, many of the current drawbacks in the governance structure might be solved or at least improved: The liability issues would be delegated away from the platforms to a third party that

86. Dijck, 51.

87. Pouwelse et al.

88. Gillespie, "Governance Of and By Platforms," 16.

89. Roberts, "Content Moderation," 1–4; Roberts, "Commercial Content Moderation."

90. Roberts, "Commercial Content Moderation: Digital Laborers' Dirty Work," 2.

91. Keller.

92. See also Langvardt; Klonick, "Facebook v. Sullivan," 16 on Facebook's dilemna between protecting users and protecting free expression.

93. Great Britain and Media and Sport Department for Culture, 46: consulting on independent review or resolution mechanisms by designated bodies; similarly: Article 19.

would be authorized to not only monitor, but also and foremost to moderate content. Its decisions might take effect across platforms and different regions, while at the same time adhering to minimum principles and considering human rights directly. Such an approach has—to a certain degree—been recognized by some platforms and currently leads to a new form of self-regulation.[94] For instance, Facebook has establishing an oversight body that reviews appeal cases that ask fundamental questions of moderating UGC.[95] According to its charter, Facebook's Oversight Board is independent from Facebook, both financially and substantively, a step that has been analyzed as a form of "enhanced self-regulation."[96] In content moderation matters, the Oversight Board is expected to decide on cases that Facebook and users cannot agree upon (categorized as "significant and difficult"), and that were selected by the Board's "selection committee."[97] The Facebook Oversight Board has decided upon first cases in 2021,[98] with follow-up decisions and policy changes expected soon after. The impact of the Oversight Board's activities remains to be seen but its creation,[99] the wider context of regulatory evasion,[100] and first decisions have already been scrutinized.[101]

Facebook is not the only social media platform undergoing change: Twitter, for example, allows its users to "appeal an account suspension or locked account" and YouTube, similarly, introduced the "right to appeal Community Guidelines strikes." Large social media platforms have generally committed themselves, for example, to publishing transparency reports, to allowing the individual right of contesting a take-down decision—moving closer to what is usually common for state actors, that is, administrative law.[102] There is still much research to be done in this area as this new type of corporate self-regulation blends in mechanisms of administrative law, but on a voluntary basis, and with limited verifiability and accountability. The advantages of such self-regulatory measures from the perspective of the users remain unclear, though.

---

94. Gorwa, 26–28.

95. Zuckerberg.

96. Medzini.

97. Facebook.

98. https://www.oversightboard.com/decision/.

99. Klonick and Kadri; Douek, "Facebook's New 'Supreme Court."; Klonick, "The Facebook Oversight Board."

100. Zalnieriute, 139–53.

101. Douek, "Facebook's New 'Supreme Court."

102. Heldt, "Let's Meet Halfway."

When distilling the abovementioned and pondering about new governance structures, it appears that moderating UGC within certain legal boundaries of private and public law would benefit from cooperation.[103] It should not be thought of as a one-to-one cooperation but rather a participatory model that would involve all actors but limited to the specific task of content moderation (i.e., preferably not Internet governance in general or other overall topics). In order to achieve the goal of effective content moderation and still shutting out a power imbalance within the triangle constellation, each actor could contribute her assets and withhold from exercising too much power unilaterally. This might also protect from the unwanted effect of proactive over-removal when intermediary liability is no longer the central resource to control.[104] Other intrinsic motifs of platforms would not be applicable either. If a third party (or fourth if counting in the user) were to monitor content moderation with its own discretionary power, it could (according to this model) be bound to societal and public interests or the common good, putting the governance of public communication back in the hands of society in the first place. Instead of "private public accountability,"[105] this approach would put content moderation decisions rather back in the realm of public accountability. The independent bodies would act as competent third parties for both the platforms and the end users, implementing agreed societal values into the platform governance structure. At the same time, this approach would not completely restrict contractual freedoms since platforms could still apply their own terms and conditions on the top of legal provisions enforced by these new institutions. Although we see an independent body as a way out of the issue described until here, the proposed governance structure also brings up new questions with regards to power structures and the true advantages of such a solution. One needs to set high expectations and standards in order to achieve an improvement of the current situation without simply "relocating" the challenges encountered to a new actor.

Involving an additional party could be a way out of the current dilemma that we either privatize the ordering of public communication or have significant state interference in cross-border communication with potential overspills in the direction of censorship. The twist is that the new player

103. Gorwa, 9, 13; Fay, 29.

104. Stalla-Bourdillon and Thorburn, 24.

105. Douek, "Self-Regulation of Content Moderation."; Douek, "Verified Accountability: Self-Regulation."

would be neither a state body, nor a platform provider or organizationally related to it; it should be legally established as a third kind of actor, empowered to decide on content moderation first instead of platforms, ensuring that content-related decisions remain in the realm of society and its self-understanding regarding public communication: a type of outsourced content moderation body, but based in human rights and public discourse instead of economic motives. This fourth player would decide on behalf of society, for example, on the basis of United Nations' Guiding Principles on Business and Human Rights.[106] This new governance perspective is not to be mistaken with the concept of information fiduciaries as developed by Balkin, which stipulates higher duties of care for intermediaries with regards to their users' data.[107] Potentially, the same goal could be achieved but the means are fundamentally different.

If these preconditions were given, the positive effects of having a neutral and independent decision-making party might show advantages in automated content decisions at scale, too. So far, there is no clear advantage for freedom of speech in using automated, usually artificial intelligence (AI)-driven systems to moderate content because, until now, those systems are not capable of making more finely nuanced decisions than humans. In copyright, for example, context-blind AI systems cannot recognize fair use or other exceptions to the protection of rights holders such as satire or parody.[108] When it comes to hate speech, AI systems are still struggling with accurate text classification and contextualization.[109] This bears significant risks for individual rights and there is a high probability of growing threats with more sophisticated technology when moderating content.[110] When it comes to partly or totally autonomous systems such as self-learning algorithms or deep learning, there is no reproducibility as there is with deterministic models. The inherent characteristic of most neural network–based algorithms being a black box makes its deployment in content moderation a delicate issue.[111] The additional layer of uncertainty raises questions as to the liability or responsibility of the actor implementing the proposed

106. As proposed in a report commissioned by Facebook and made by BSR on the basis of the UNGP, Allison-Hope, Lee, and Lovatt.

107. Balkin, "Information Fiduciaries," 52.

108. Perel and Elkin-Koren, "Black Box Tinkering: Beyond Disclosure," 13.

109. Duarte, Llanso, and Loup, 106.

110. Aken et al., 7 f.; Heldt, "Upload-Filters: Bypassing Classical Concepts of Censorship?" 65.

111. Perel (Filmar) and Elkin-Koren, "Separation of Functions for AI," 40.

system. All in all, the shift toward a system run by an AI would deserve consideration but definitely needs additional reflection.[112] Competent third parties would seem structurally advantageous here in general, as they have no vested interest in business secrets around AI-based automated content moderation technologies. The development of more human rights–sensitive AI decision models could not only be open-sourced, they could also be controlled and optimized in the most reflexive way, since neither company secrecy nor economic considerations would make these procedures to necessarily be opaque: As institutions without economic or political business logics, they could implement highly and openly tested models and undergo systematic external monitoring and evaluation measures. Moreover, the optimized models could be deployed across platforms, helping to overcome current intransparency in automated decision-making in platforms' content moderation.

Requirements and Procedural Arrangements for Competent Third Parties

To benefit from the advantages of a third-party structure like we described earlier, social media platforms would need to outsource their content moderation activities when it comes to compliance with the law, not only for the review of contested take-down decisions. A privilege in terms of liability can only be granted to those platforms that choose to opt in some sort of a safe harbor.[113] Opting in such a model would not be mandatory by law but if platforms want to dispose of their liability for unlawful content it would require a real transfer of competency as to the basic legal standards. It would still be at their discretion to maintain higher standards as long as it does not come at the disproportionate expense of freedom of expression and information.

In the following, we will outline the advantages and disadvantages of a trusted independent body from a governance structure perspective, as well as the requirements and procedural arrangements. Most of the suggestions surely will have to be discussed more thoroughly. We have already mentioned several reasons why adding a neutral body to the existing speech triangle would be beneficial and helpful for the unsolved problems of content moderation. The most obvious advantage is the responsibility shift

112. See also Gorwa, Binns, and Katzenbach.

113. Balkin, "Free Speech Is a Triangle," 2046.

from social media platforms to a new player when it comes to content moderation. Content moderation is a moving target, a field that regularly requires new policy decisions while bearing in mind the legal perspective. By establishing a new actor in the online infrastructure, which would not be affiliated with single social media platforms, cross-platform decision practices as well as coherent case law would be made possible if various platforms adhere to its decisions. Although it would be naïve to expect that the decisions would be applicable on a global scale, they could provide guidance across platforms and within cultures. This aspect should also be considered on a global scale and be taken in consideration when distributing responsibilities. In other words, there needs to be an equal distribution of power on a regional level, more user-centered and mindful of their respective societal values. It will probably be necessary to install several of those external bodies or internal units depending on the regional scope of applicable legal frameworks to consider different interpretations of freedom of speech; however, the body could ensure that a minimum standard regarding the consideration of human rights would be upheld. This could also translate into a cost advantage, a centralization of expertise, and higher legal certainty for both the platforms and the users. The competent third party could serve as a reliable contact point for both state authorities and platform providers ("content moderation intermediary"), setting minimum standards and best practices across platforms. Moreover, it would have the advantage of establishing sustainable procedures while simultaneously providing an agile form of reaction to new problems arising without the obligation to adhere to business logics or finding political majorities in decision-making. It would be the central node for affected users, granting them globally exercisable rights toward a body that is neither a contracting party nor a bottleneck to information and communication.

On the flip side, one could criticize that centralizing all decision-making into one powerful actor would signify handing over too much power over online communication. One would need to spell out safeguards against this potential shift in power over individual and public communication. Moreover, the platforms would lose control over the content published on their own applications and might fear so-called "sovereignty costs." In Abbott and Snidal's concept, sovereignty costs are the highest when the delegation of power by a state leads to a self-limitation of its legislative power.[114] What has so far been analyzed for states (making commitments

114. Ibid., 437.

and delegating sovereignty to an organization on a higher, supranational level) could now be applied to nonstate actors with important market power: transferring the task of content moderation would confine platforms substantively. Considering that we are looking at content generated by users, there is no real ownership of UGC by the platforms but it would still transform them into mere distributors of undisputed (or even desired) legal content if they lost control over content moderation. Additionally, potential conflicts between the new body's decisions and the respective terms of service of social media platforms may arise. If the restrictions are potentially endangering their business model or simply making it (much) less attractive, social media platforms are not likely to adopt the proposed model. However, the advantages could outperform the disadvantages considering that the latter are by no means insurmountable obstacles.

Establishing such external bodies would only come with regulatory advantages in case of safeguarding specific aspects: It would need to guarantee independence from both state and platforms, and offer transparency and due process in a way in which private actors are generally not bound to. By doing so, it would combine the advantages of the actors who currently are in an advantageous power position over online speech in comparison to users, that is, the legislator and the platforms. This goes hand in hand with their resemblance with state actors: the more power private actors gain, the more they tend to be assimilated with the state or state-like structures.[115] In the context of online speech, large platforms have been called "custodians,"[116] "new governors,"[117] "deciders,"[118] and more designations that show the close link between power and societal expectations toward state-like actors.

One requirement here, with regard to the institutional differentiation, must be to draw clear lines between the actors involved. To avoid the allegation of censorship, the competent third body would need to be independent enough from state authority, and, similarly, from corporations to circumvent allegations of lobbying or revolving doors. In an effort to combine the best of all worlds, the competent third party would be user participation- and common-good oriented and particularly toward freedom of speech and information. Its decisions would be

115. Heldt, "Let's Meet Halfway," 358 f.
116. Gillespie, Custodians of the Internet.
117. Klonick, "The New Governors."
118. Rosen, 1525.

accessible for judicial review, but the external body would not be liable. Hence, it would not undermine the judiciary but rather intercept cases that can be decided by internal content moderators, leaving it up to the users or platforms to decide whether they wish to file a lawsuit in specific cases. Its decisions would be subject to judicial review as common litigation, or the procedure with the competent third party could be exhaustive if it is considered as some form of alternative dispute resolution. In light of the international context and the cross-border circumstances, it might be preferable to traditional litigation and national judicial review. However, such alternative dispute resolution system would require absolute transparency to avoid falling in the same opacity trap as before. Concerning accountability, the alternative governance structure would result in a somewhat radical shift: Platform providers would "outsource" their responsibilities regarding illegal user content, solving the current liability dilemma that sets incentives either to look away or to block also borderline content. A central purpose of establishing competent third parties is to shift responsibility for assessing illegal content away from the platforms toward the new content moderation bodies, mitigating the ambivalent incentive of a liability framework for platforms. However, the platform would still remain accountable for all its content moderation activities taking place on grounds of "only" the community standards, preventing the platforms from presenting themselves to the outside world as blameless and not accountable. For illegal content that is accessed by the external party once a platform submits to such a system, the platforms cannot be held liable for illegal content anymore, since the responsibility for moderating such content lies now with the third party. Like courts, these actors decide on the lawfulness of a specific piece of content. By following due process of law procedures, persons affected by the decisions could only sue the decision-making body in cases where the body exceeds its margin of appreciation in an apparent manner. (Like before, such proceedings would not hinder a person infringed in his or her rights to sue before a civil court against the person who posted the content in question.)

Another important requirement is the internal governance of such a competent third party: what has been described in this article for the actors is also valid for the new player's structure. Many of our arguments can be attributed to the principle of separation of powers, that is, a separation of legislative, executive, and judiciary. The internal structure should reflect this principle by sharing corresponding tasks and

responsibilities. Developing content moderation rules shall, therefore, be separated from their application and opposition proceedings. The legislative is the norm-making power which, at first, remains with states (national legislation) and platforms (terms and conditions) but could, at a later stage, be further developed in a process involving all parties.[119] What about the precedent of decisions after an appeal? They could develop normative effects in addition to the hitherto corpus of legislation. Case-by-case decisions regarding individual cases would fall under the judicial review of the competent third party, including a higher level of jurisdiction to ensure impartial decision-making. After that, each case needs to be enforced, whether it is a first stage decision to sanction UGC or an appeal decision. This part would remain within the responsibility of the platforms since, as described earlier, enforcement is the main feature of their "sphere of efficacy." The executive power, in other words, the operational side of social media would, in any case, be the platforms' own. We were aiming for a conceptual "Gedankenexperiment" and would welcome more discussion when it comes to practical details and subsequent issues.

Outlook

As a result, we are not moving away from the classic triangle structure of online speech but we offer a new perspective on it by establishing a new form of body in the very center of the triangle. It would be positioned in between the current spheres of efficacy, pushing a new model based on cooperation in- and outside the new body. We do not see it as a mere "breaker" between the state and platforms. In fact, we position it at the center of the triangle because it would assume responsibilities of both the state and the platforms to the advantage of the individual whose rights would no longer be fragmented between freedom of expression and contractual rights as a consumer. It would reduce the users' lack of control that we have identified when laying out the spheres of efficacy,[120] because instead of being governed by many, content control will be exercised by one centralized body. At the same time, users will be able

119. The question of how to further develop rules once the here outlined model is introduced is one of its own and would go beyond the scope of this article.

120. See supra chap. 4.

to invoke their right to freedom of expression and due process in an unprecedented way in content moderation. This competent third party should be understood as a chance to recouple content moderation with society while at the same time relieving the platforms from their liability dilemma when it comes to moderating UGC. A risk of this shift in responsibility to such a competent third-party model is that it could be perceived as a mere strategy of blame shifting. To avoid such scenario, we ought to conceptualize it as a possibility to organize a better distribution of power (rulemaking, judgment, and enforcement) and to reach decisions that apply across multiple platforms, establishing a human rights–based ground level of content moderation. Simultaneously, it could contribute to protecting legitimate speech, thus limiting the over-removal of UGC. This would apply to decision-making at scale, too, especially in the realm of (semi-) automated systems. If content moderation is outsourced to an independent competent third party, its activities could include gathering data, training AI, and facilitating the development of a system that would eventually become a trusted content moderation AI. The independent body could be responsible for the data processing aiming at optimizing the constant work in progress of content moderation. The data would be used in order to guarantee content moderation standards: intelligent systems could then apply, at least to support human moderators. Ideally, such an approach will be better suited to implement transparent and evaluated procedures and technologies. The main advantage is that, according to our proposal, the third party would not have to follow the platforms' business logic, that is, leaving behind the necessity of overlooking potentially harmful and prioritizing contentious content. The alternative approach here is to decouple human rights–relevant decision-making structures from ad-based business models; by doing so, we will be able (again) to discuss about forms of wanted and unwanted communications from a societal and not from a business perspective.

Conflict of Interest

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that there is no conflict of interest. All authors have agreed to the submission and the article is not currently under review at a different journal.

BIBLIOGRAPHY

Abbott, Kenneth W., and Duncan Snidal. "Hard and Soft Law in International Governance." *International Organization* 54, no. 3 (2000): 421–56. https://doi.org/10.1162/002081800551280.

———. "The Governance Triangle: Regulatory Standards Institutions and the Shadow of the State." In *The Politics of Global Regulation*, edited by Walter Mattli and Ngaire Woods, 44–88. Princeton: Princeton University Press, 2009. https://doi.org/10.1515/9781400830732.44.

Aken, Betty van, Julian Risch, Ralf Krestel, and Alexander Löser. "Challenges for Toxic Comment Classification: An In-Depth Error Analysis." *ArXiv:1809.07572 [Cs]*, September 20, 2018. http://arxiv.org/abs/1809.07572.

Allison-Hope, Dunstan, Michaela Lee, and Joanna Lovatt. *Human Rights Review: Facebook Oversight Board*. New York, United States: BSR | Business for Social Responsibility, December 2019. https://www.bsr.org/reports/BSR_Facebook_Oversight_Board.pdf.

Article 19. "The Social Media Councils: Consultation Paper." June 2019. https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

Balkin, Jack M. "Free Speech Is a Triangle." *Columbia Law Review* 118, no. 7 (May 28, 2018): 2011–56.

———. "Information Fiduciaries and the First Amendment." *UC Davis Law Review* 49 (2016): 52.

———. "Old School/New School Speech Regulation." *Harvard Law Review* 127, no. 8 (June 2014): 2296–2342.

Baron, David P. "Design of Regulatory Mechanisms and Institutions." In *Handbook of Industrial Organization*, edited by Richard Schmalensee and Robert D. Willig, 1349. Handbooks in Economics 10. Amsterdam: North-Holland, 1989.

Batorski, Dominik, and Ilona Grzywińska. "Three Dimensions of the Public Sphere on Facebook." *Information, Communication & Society* 21, no. 3 (March 4, 2018): 356–74. https://doi.org/10.1080/1369118X.2017.1281329.

Belli, Luca, and Jamila Venturini. "Private Ordering and the Rise of Terms of Service as Cyber-Regulation." *Internet Policy Review* 5, no. 4 (December 29, 2016). https://doi.org/10.14763/2016.4.441.

Birnhack, Michael D., and Niva Elkin-Koren. "The Invisible Handshake: The Reemergence of the State in the Digital Environment." *Virginia Journal of Law and Technology* 8, no. 6 (2003): 1–56.

Black, J. "Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World." *Current Legal Problems* 54, no. 1 (January 1, 103): 2001–46. https://doi.org/10.1093/clp/54.1.103.

Borghol, Youmna, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. "The Untold Story of the Clones: Content-Agnostic Factors That Impact YouTube Video Popularity." In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, 1186. Beijing: ACM Press, 2012. https://doi.org/10.1145/2339530.2339717.

Bradley, Charles, and Richard Wingfield. *A Rights-Respecting Model of Online Content Regulation by Platforms*. London: Global Partners Digital, March 2018.

Buni, Catherine, and Soraya Chemaly. "The Secret Rules of the Internet." *The Verge*, April 13, 2016. https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech.

Cammaerts, Bart, and Robin Mansell. "Digital Platform Policy and Regulation: Toward a Radical Democratic Turn." *International Journal of Communication* 14 (January 1, 2020): 135–54.

Cannataci, Joseph A., and Jeanne Pia Mifsud Bonnici. "Can Self-Regulation Satisfy the Transnational Requisite of Successful Internet Regulation?" *International Review of Law, Computers & Technology* 17, no. 1 (March 2003): 51–61. https://doi.org/10.1080/13600860320000063110.

Citron, Danielle Keats. "Extremist Speech, Compelled Conformity, and Censorship Creep." *Notre Dame Law Review* 93, no. 3 (2017): 1035.

Citron, Danielle, and Jurecic Quinta. "Platform Justice: Content Moderation at an Inflection Point." Aegis Series. Stanford: Hoover Institution, May 9, 2018. https://www.hoover.org/research/platform-justice.

Collins, Brian J., Jose Marichal, and Richard Neve. "The Social Media Commons: Public Sphere, Agonism, and Algorithmic Obligation." *Journal of Information Technology & Politics* 17, no. 4 (April 2, 2020): 1–17. https://doi.org/10.1080/19331681.2020.1742266.

Council of Europe. "Recommendation CM/Rec(2012) of the Committee of Ministers to Member States on the Protection of Human Rights with Regard to Social Networking Services." 2012. https://wcd.coe.int/ViewDoc.jsp?id=1929453.

Crawford, Kate, and Tarleton Gillespie. "What Is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint." *New Media & Society* 18, no. 3 (March 1, 2016): 410–28. https://doi.org/10.1177/1461444814543163.

Crockett, M. J. "Moral Outrage in the Digital Age." *Nature Human Behaviour* 1, no. 11 (November 2017): 769–71. https://doi.org/10.1038/s41562-017-0213-3.

De Gregorio, Giovanni. "Democratising Online Content Moderation: A Constitutional Framework." *Computer Law & Security Review* 36 (April 2020): 105374. https://doi.org/10.1016/j.clsr.2019.105374.

Deibert, Ronald J. "Three Painful Truths About Social Media." *Journal of Democracy* 30, no. 1 (2019): 25–39. https://doi.org/10.1353/jod.2019.0002.

Dijck, José van. "Users Like You? Theorizing Agency in User-Generated Content." *Media, Culture & Society* 31, no. 1 (January 2009): 41–58. https://doi.org/10.1177/0163443708098245.

Dijck, José Van, and Thomas Poell. "Understanding Social Media Logic." *Media and Communication* 1, no. 1 (August 12, 2013): 2–14. https://doi.org/10.12924/mac2013.01010002.

Douek, Evelyn. "Facebook's New 'Supreme Court' Could Revolutionize Online Speech." *Lawfare* (blog), November 19, 2018. https://www.lawfareblog.com/facebooks-new-supreme-court-could-revolutionize-online-speech.

———. "Self-Regulation of Content Moderation as an Answer to the Special Problems of Speech Regulation." Aegis Series. National Security, Technology, and Law. Stanford: Hoover Institution, 2019.

———. "Verified Accountability: Self-Regulation of Content Moderation as an Answer to the Special Problems of Speech Regulation." *Lawfare* (blog), September 18, 2019. https://www.lawfareblog.com/verified-accountability-self-regulation-content-moderation-answer-special-problems-speech-0.

Dreyer, Stephan. "Locating a Regulator in the Governance Structure: A Theoretical Framework for the Operationalization of Independence." In *The Independence of the Media and Its Regulatory Agencies: Shedding New Light on Formal and Actual Independence against the National Context*, edited by Wolfgang Schulz, Peggy Valcke, and Kristina Irion, 111. Bristol: Intellect, 2013.

Dreyer, Stephan, and Lennart Ziebarth. "Participatory Transparency in Social Media Governance: Combining Two Good Practices." *Journal of Information Policy* 4 (2014): 529. https://doi.org/10.5325/jinfopoli.4.2014.0529.

Duarte, Natasha, Emma Llanso, and Anna Loup. "Mixed Messages? The Limits of Automated Social Media Content Analysis." In *FAT*, 106, 2018.

Echikson, William, and Olivia Knodt. "Germany's NetzDG: A Key Test for Combatting Online Hate." Research report. Thinking Ahead of Europe. Brussels, Belgium: CEPS, November 2018. www.ceps.eu.

Elkin-Koren, Niva, and Eldar Haber. "Governance by Proxy: Cyber Challenges to Civil Liberties." *Brooklyn Law Review* 82 (2017 2016): 105–62.

EU Commission. "Protecting European Democracy from Interference and Manipulation – European Democracy Action Plan." Accessed October 30, 2020. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12506-European-Democracy-Action-Plan.

European Commission. *Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, 12th September 2018, COM(2018) 640 final.* https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-regulation-640_en.pdf.

Facebook. "Draft Charter: An Oversight Board for Content Decisions." *Facebook*, January 28, 2019. https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-2.pdf.

Fagan, Frank. "Systemic Social Media Regulation." *Duke Law & Technology Review* 16 (2018 2017): 393.

Fay, Robert. "Digital Platforms Require a Global Governance Framework." Essay. Models for Platform Governance. Waterloo: Centre for International Governance Innovation. Accessed February 24, 2020. https://www.cigionline.org/articles/digital-platforms-require-global-governance-framework.

Gasser, Urs, and Wolfgang Schulz. "Governance of Online Intermediaries: Observations From A Series Of National Case Studies." *Berkman Center Research Publication No. 2015-5*, 2015. https://cyber.law.harvard.edu/publications/2015/online_intermediaries.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven; London: Yale University Press, 2018.

———. "Governance of and by Platforms." In *Sage Handbook of Social Media*, edited by Jean Burgess, Alice Marwick, and Thomas Poell, 254–78. Los Angeles: Sage, 2017.

———. "The Platform Metaphor, Revisited." *HIIG* (blog), August 24, 2017. https://www.hiig.de/en/the-platform-metaphor-revisited/.

———. "The Politics of 'Platforms.'" *New Media & Society* 12, no. 3 (May 2010): 347–64. https://doi.org/10.1177/1461444809342738.

Gorwa, Robert. "The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content." *Internet Policy Review* 8, no. 2 (June 30, 2019). https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content.

Gorwa, Robert, Reuben Binns, and Christian Katzenbach. "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance." *Big Data & Society* 7, no. 1 (January 2020): 1–15. https://doi.org/10.1177/2053951719897945.

Great Britain Media and Sport Department for Culture. *Online Harms White Paper*, 2019. Presented to Parliament by the Secretary of State for Digital, Culture, Media & Sport and the Secretary of State for the Home Department by Command of Her Majesty, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf.

C

Greis, Miriam, Florian Alt, Niels Henze, and Nemanja Memarovic. "I Can Wait a Minute: Uncovering the Optimal Delay Time for Pre-Moderated User-Generated Content on Public Displays." In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, 1435–38, edited by Matt Jones and Philippe Palanque. Toronto, ON, Canada: ACM Press, 2014. https://doi.org/10.1145/2556288.2557186.

Grimmelmann, James. "The Virtues of Moderation." *Yale Journal of Law & Technology* 17, no. 1 (2015): 43–109.

Hart, H. L. A. *The Concept of Law*. 3rd ed. Clarendon Law Series. Oxford: Oxford University Press, 2012.

Helberger, Natali, Jo Pierson, and Thomas Poell. "Governing Online Platforms: From Contested to Cooperative Responsibility." *The Information Society* 34, no. 1 (January 2018): 1–14. https://doi.org/10.1080/01972243.2017.1391913.

Helberger, Natali, Katharina Kleinen-von Königslöw, and Rob van der Noll. "Regulating the New Information Intermediaries as Gatekeepers of Information Diversity." *Info* 17, no. 6 (2015): 50–71.

Heldt, Amélie. "Let's Meet Halfway: Sharing New Responsibilities in a Digital Age." *Journal of Information Policy* 9, no. Special Issue: Government and Corporate Policies for Social Media (2019): 336–69. https://doi.org/10.5325/jinfopoli.9.2019.0336.

Heldt, Amélie P. "Von der Schwierigkeit, Fake News zu regulieren: Frankreichs Gesetzgebung gegen die Verbreitung von Falschnachrichten im Wahlkampf." *bpb.de (blog)*, May 2, 2019. http://www.bpb.de/gesellschaft/digitales/digitale-desinformation/290529/frankreichs-gesetzgebung-gegen-die-verbreitung-von-falschnachrichten.

Heldt, Amélie Pia. "Upload-Filters: Bypassing Classical Concepts of Censorship?" *JIPITEC* 10, no. 1 (May 5, 2019): para 1.

Hirsch, Dennis D. "The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation?" *Seattle University Law Review* 34 (2010): 439.

Hofmann, Jeanette, Christian Katzenbach, and Kirsten Gollatz. "Between Coordination and Regulation: Finding the Governance in Internet Governance." *New Media & Society* 19, no. 9 (September 2017): 1406–23. https://doi.org/10.1177/1461444816639975.

Katzenbach, Christian, Jeanette Hofmann, Kirsten Gollatz, Francesca Musiani, Dmitry Epstein, Laura DeNardis, Andrea Hackl, and Jean-François Blanchette. "Doing Internet Governance: STS-Informed Perspectives on Ordering the Net." *First Monday*, *AoIR Selected Papers of Internet Research*, 5, no. 0 (2015). https://firstmonday.org/ojs/index.php/spir/article/view/8564.

Kaye, David. "Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression." *Human Rights Council. United Nations General Assembly*, April 6, 2018. https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/OpinionIndex.aspx.

Keller, Daphne. "Who Do You Sue? State and Platform Hybrid Power Over Online Speech." Aegis Series. National Security, Technology, and Law. Stanford: Hoover Institution, January 30, 2019. https://www.hoover.org/research/who-do-you-sue.

Keller, Daphne, and Paddy Leerssen. "Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation." In *Social Media and Democracy: The State of the Field and Prospects for Reform*, edited by Nathaniel Persily and Joshua A. Tucker, 46. Cambridge, UK: Cambridge University Press, 2020.

Kettemann, Matthias C., and Wolfgang Schulz. "Setting Rules for 2,7 Billion – A (First) Look into Facebook's Norm-Making System: Results of a Pilot Study." Works in Progress #1. Hamburg, Germany: Leibniz-Institute for Media Research, January 2020.

Kinstler, Linda. "Can Germany Fix Facebook?" *The Atlantic*, November 2, 2017. https://www.theatlantic.com/international/archive/2017/11/germany-facebook/543258/.

Klonick, Kate. "Facebook v. Sullivan." *Knight First Amendment Institute: Emerging Threats Series* (blog), October 2018.

———. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression." *Yale Law Journal* 129, no. 8 (2020): 2232–2605.

———. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131, no. 6 (2018): 1598–1669.

Klonick, Kate, and Thomas Kadri. "How to Make Facebook's 'Supreme Court' Work." *The New York Times*, November 18, 2018, sec. Opinion. https://www.nytimes.com/2018/11/17/opinion/facebook-supreme-court-speech.html.

Kornhauser, Lewis A. "Governance Structures, Legal Systems, and the Concept of Law." *Chicago-Kent Law Review* 79, no. 2 (2004): 355–81.

Kreimer, Seth F. "Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link." *University of Pennsylvania Law Review*, Faculty Scholarship, Paper 127, 155 (2006): 11.

Langvardt, Kyle. "Regulating Online Content Moderation." *Georgetown Law Journal* 106, no. 5 (2018): 1353–88. https://doi.org/10.2139/ssrn.3024739.

Latzer, Michael, Natascha Just, and Florian Saurwein. "Self- and Co-Regulation: Evidence, Legitimacy and Governance Choice." In *Routledge Handbook of Media Law*, edited by Monroe Edwin Price, Stefaan Verhulst, and Libby Morgan, 373–97. Routledge Handbooks. London: Routledge, 2013.

Lessig, Lawrence. *Code: Version 2.0*. 2nd ed. New York: Basic Books, 2006.

Loo, Rory Van. "Rise of the Digital Regulator." *Duke Law Journal* 66 (2017): 1267.

Majone, Giandomenico. "Regulation and Its Modes." In *Regulating Europe*, edited by Giandomenico Majone, 9–27. European Public Policy. London: Routledge, 1996.

Malhotra, Neil, Benoît Monin, and Michael Tomz. "Does Private Regulation Preempt Public Regulation?" *American Political Science Review* 113, no. 1 (February 2019): 19–37. https://doi.org/10.1017/S0003055418000679.

Marantz, Andrew. "Facebook and the 'Free Speech' Excuse." *The New Yorker*, October 31, 2019. https://www.newyorker.com/news/daily-comment/facebook-and-the-free-speech-excuse.

Medzini, Rotem. "Enhanced Self-Regulation: The Case of Facebook's Content Governance." *New Media & Society*, February 1, 2021, 1461444821989352. https://doi.org/10.1177/1461444821989352.

Myers West, Sarah. "Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms." *New Media & Society* 20, no. 11 (November 2018): 4366–83. https://doi.org/10.1177/1461444818773059.

Napoli, Philip M. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." *Telecommunications Policy*, Special Issue on the Governance of Social Media, 39, no. 9 (October 1, 2015): 751–60. https://doi.org/10.1016/j.telpol.2014.12.003.

Peeters, Kris. "Digital Services Act and Fundamental Rights Issues Posed." *Brussels: European Parliament*, October 20, 2020. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0274_EN.pdf.

Perel, Maayan, and Niva Elkin-Koren. "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement." *Florida Law Review* 69, no. 1 (2017): 181–221.

———. "Separation of Functions for AI: Restraining Speech Regulation by Online Platforms," August 19, 2019. https://papers.ssrn.com/abstract=3439261.

Pouwelse, Johan A., Pawel Garbacki, Dick Epema, and Henk Sips. "Pirates and Samaritans: A Decade of Measurements on Peer Production and Their Implications for Net Neutrality and Copyright." *Telecommunications Policy* 32, no. 11 (December 2008): 701–12. https://doi.org/10.1016/j.telpol.2008.09.004.

Puppis, Manuel. "Media Governance: A New Concept for the Analysis of Media Policy and Regulation." *Communication, Culture & Critique* 3, no. 2 (June 2010): 134–49. https://doi.org/10.1111/j.1753-9137.2010.01063.x.

Raymond, Mark, and Laura DeNardis. "Multistakeholderism: Anatomy of an Inchoate Global Institution." *International Theory* 7, no. 3 (November 2015): 572–616. https://doi.org/10.1017/S1752971915000081.

Roberts, Sarah T. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven: Yale University Press, 2019.

———. "Commercial Content Moderation: Digital Laborers' Dirty Work." *FIMS Wester University Media Studies Publications*, 2016, 1–12.

———. "Content Moderation." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, 1–4. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-32001-4_44-1.

Rosen, Jeffrey. "The Deciders: The Future of Privacy and Free Speech in the Age of Facebook and Google." *Fordham Law Review* 80 (2011): 1525.

Ruckenstein, Minna, and Linda Lisa Maria Turunen. "Re-Humanizing the Platform: Content Moderators and the Logic of Care." *New Media & Society*, September 19, 2019, 1461444819875990. https://doi.org/10.1177/1461444819875990.

Schulz, Wolfgang, and Thorsten Held. *Regulated Self-Regulation as a Form of Modern Government: An Analysis of Case Studies from Media and Telecommunications Law*. Eastleigh, UK: Published by John Libbey for University of Luton Press, 2004.

———. "Study on Co-Regulation Measures in the Media Sector." *Final Report. Study for the European Commission, Directorate Information Society and Media Unit A1 Audiovisual and Media Policies*. Brussels, Belgium: European Commission, 2006. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/cd368d1fee0e-0cee4d50061f335e5629184612 45.pdf.

Schwarcz, Steven L. "Private Ordering." *Northwestern University Law Review* 97, no. 1 (2002): 319–50.

Stalla-Bourdillon, Sophie, and Robert Thorburn. "The Scandal of Intermediary: Acknowledging the Both/and Dispensation for Regulating Hybrid Actors." In *Fundamental Rights Protection Online: The Future Regulation of Intermediaries*, edited by Bilyana Petkova and Tuomas Ojanen, 25. Cheltenham: Edward Elgar, 2019.

Suzor, Nicolas P. *Lawless: The Secret Rules That Govern Our Digital Lives*. Cambridge; New York: Cambridge University Press, 2019.

Szabo, Gabor, and Bernardo A. Huberman. "Predicting the Popularity of Online Content." *ArXiv:0811.0405 [Physics]*, November 4, 2008. http://arxiv.org/abs/0811.0405.

Tambini, Damian. "Social Media Power and Election Legitimacy." In *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*, edited by Martin Moore and Damian Tambini, 265–93. New York: Oxford University Press, 2018.

Trute, Hans-Heinrich, Doris Kühlers, and Arne Pilniok. "Rechtswissenschaftliche Perspektiven." In *Handbuch Governance – Theoretische Grundlagen und Empirische Anwendungsfelder*, edited by Arthur Benz, Susanne Lütz, Uwe Schimank, and Georg Simonis, 240–52. Wiesbaden: VS Verlag für Sozialwissenschaften, 2007. https://link-1springer-1com-10039ffj40066.emedien3.sub.uni-hamburg.de/chapter/10.1007/978-3-531-90407-8_18.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. "From Liberation to Turmoil: Social Media and Democracy." *Journal of Democracy* 28, no. 4 (October 7, 2017): 46–59. https://doi.org/10.1353/jod.2017.0064.

Tufekci, Zeynep. "How Social Media Took Us from Tahrir Square to Donald Trump." *MIT Technology Review*, October 2018.

Tushnet, Rebecca. "Power Without Responsibility: Intermediaries and the First Amendment." *George Washington Law Review* 101 (August 6, 2008): 101–33.

Weber, Rolf H., and R. Shawn Gunnarson. "A Constitutional Solution for Internet Governance." *Columbia Science & Technology Law Review* 14, no. 1 (2012): 52.

Yemini, Moran. "Free Speech for All: A Justice-Infused Theory of Speech for the Digital Ecosystem." Doctoral Dissertation, Center for Cyber Law and Policy, University of Haifa, 2017.

Zalnieriute, Monika. "'Transparency Washing' in the Digital Age: A Corporate Agenda of Procedural Fetishism." *Critical Analysis of Law* 8, no. 1 (April 2, 2021): 139–53.

Zuckerberg, Mark. "A Blueprint for Content Governance and Enforcement." *Facebook (blog)*, November 15, 2018. https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/.