Enhancing Legitimacy of Content Moderation

by **Jelizaveta Juřičková** *

Abstract: Platforms are actively developing strategies to enhance the legitimacy of their content moderation and gain acceptance and trust across diverse user groups. This paper explores one such strategy, endorsed by the EU regulator, which involves proceduralizing content moderation, with a focus on copyright enforcement as a case study.
However, the paper raises concerns regarding the efficacy of proceduralization in legitimizing content moderation, citing historical limitations in the adoption of dispute resolution mechanisms by ordinary users. In response, the paper suggests a complementary approach: integrating elements of procedural justice, based on users' perceptions of fairness, into the implementation of content moderation requirements mandated by regulators. By elucidating how procedural justice enhances legitimacy and drawing from user experiences with content moderation, the paper proposes a preliminary index of procedural justice values to be used as a metric and guidance for putting regulatory requirements into practice.

Keywords: Online Platforms; Content Moderation; Procedural Justice

Recommended citation: Jelizaveta Juřičková, Enhancing Legitimacy of Content Moderation, 15 (2024) JIPITEC 2 para 1.

A. Introduction

1 Initially, rightsholders struggled to enforce copyright against individual internet users, only to later pivot their approach by enlisting the assistance of online intermediaries, including online platforms, as "innocent bystanders".[1] Now, we find ourselves in a time where the dynamics have changed once more, as platforms are mandated to take a proactive role in enforcing copyright, as evident in CDSM[2] and DSA.[3] However, this newfound responsibility has left

* Ph.D. candidate at the Institute of Law and Technology, Masaryk University. This article was written at Masaryk University as part of the project n. MUNI/A/1529/2023 - Právo a technologie XII.

1 Martin Husovec, *Injunctions against Intermediaries in the European Union: Accountable but Not Liable?* (Cambridge University Press 2017) <https://www.cambridge.org/core/books/injunctions-against-intermediaries-in-the-european-union/A42D5F859EF35FAF33C2FC4EB65A6AAA>.

2 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

3 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

platforms ill at ease, as they face increased scrutiny from various stakeholders: the public, including the platform users, creative industries, and academia.[4] In response, platforms are devising strategies to legitimize their content moderation efforts, seeking acceptance and trust from these diverse groups.

2 One such strategy involves proceduralization of content moderation. This approach has also been embraced by EU regulator as a means to bring structure and accountability to the process. While this is a positive step, we must ask ourselves, is it enough to win over the general public? My argument suggests that it might not be sufficient, particularly considering that the dispute resolution mechanisms, the very vehicles of proceduralization approach, have historically seen limited adoption by ordinary users.[5]

3 Considering this, a complementary approach is proposed: focusing on procedural justice in the psychological sense when implementing content moderation requirements imposed by the regulator. This approach goes hand in hand with proceduralization, complementing it while emphasizing a different aspect. By prioritizing procedural justice, platforms can foster a notion of fair content moderation among users, thereby favourably changing their perception of its legitimacy. This emphasis on procedural justice could bridge the gap between platforms' efforts to enforce copyright and the acceptance and understanding of these measures by the broader public.

4 Section 2 introduces the proceduralization trend in content moderation and its role as a platform governance legitimation strategy. Section 3 provides examples of proceduralization within platform initiatives, focusing on copyright content moderation as a case study. Section 4 analyzes the EU-level regulatory framework governing copyright content moderation and sheds light on the limits of the proceduralization approach embedded in the regulatory framework. The paper posits that proceduralization primarily promotes legality-based legitimacy while neglecting sociological legitimacy. Furthermore, it is maintained that dispute resolution mechanisms as pivotal components of proceduralization, rely on adoption by users, which historically tends to be low. To address the limitations discussed in Section 4, Section 5 proposes a complementary legitimation strategy. This approach involves integrating elements of procedural justice, as derived empirically from psychological research, into the practical implementation of the regulatory framework by both platforms and dispute resolution bodies. These elements have been shown to influence sociological legitimacy and complement formal legality. The paper further discusses how EU regulator can incentivize platforms and dispute resolution bodies to adhere to this strategy. Section 6 we summarizes key findings and insights from the preceding sections.

B. Proceduralization Approach in Content Moderation

5 Proceduralization of content moderation refers to the process of establishing explicit rules, procedures, and standards for content moderation on online platforms. It involves making the content moderation process more structured and systematic, akin to legal or judicial systems.[6] Proceduralization comprises the following aspects — due process, quality of decisions and transparency.

6 Incorporation of safeguards of due process ensures that users whose content is being moderated have certain rights and protections.[7] This might include the right to report a piece of content that breaches the user's rights, the right to be notified about the action undertaken towards the content and be provided with justification and the right to appeal the decision. Internal mechanisms for reviewing the appeal by the platform present a particularly fruitful ground for implementation of due process features. A meaningful pendant to review by platforms are external mechanisms for settlement of disputes,

4 Taddeo and Floridi bring a comprehensive overview of the discourse regarding responsibilities of intermediaries, empirically demonstrating its evolution by evaluating relevance of the topics based on the volume of literature dedicated to each topic. Mariarosaria ed. Taddeo and Luciano ed. Floridi, *The Responsibilities of Online Service Providers* (1., Springer International Publishing AG). Chapter 2.

5 Lenka Fiala and Martin Husovec, 'Using Experimental Evidence to Improve Delegated Enforcement' (3 March 2022) <https://papers.ssrn.com/abstract=3218286> accessed 9 May 2023; Jennifer M Urban, Joe Karaganis and Brianna L Schofield, 'Notice and Takedown: Online Service Provider and Rightsholder Accounts of Everyday Practice' (2017) 64 Journal of the Copyright Society of the USA 317; Jennifer M Urban and Laura Quilter, 'Symposium Review Efficient Process or " Chilling Effects "? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act' (2000) 512 621.

6 Evelyn Douek, 'The Siren Call of Content Moderation Formalism' (10 January 2022) <https://papers.ssrn.com/abstract=4005314> accessed 21 June 2023.

7 Catalina Goanta and Pietro Ortolani, 'Unpacking Content Moderation: The Rise of Social Media Platforms as Online Civil Courts' (22 November 2021) <https://papers.ssrn.com/abstract=3969360> accessed 19 June 2023 p. 18.

promoted by the regulator.

7 The second aspect of proceduralization is raising standards for quality of decisions in content moderation. By drawing on the principles and logic used in judicial systems, consistent and coherent reasoning is applied to content decisions.[8] It involves following past decisions as precedents for current and future rulings, creating a sense of predictability.[9]

8 The third aspect, transparency, involves explaining the steps of content moderation, i.e. laying out the specific actions and procedures that content moderators follow when evaluating and handling content. It presents a *conditio sine qua non* for control of content moderation by public, academia and the regulator by offering the insight into the actual content moderation practices.

9 Formalizing content moderation has a significant potential to improve its legitimacy. The main legitimacy concepts are normative, focusing on the justification of power, sociological, which examines how the subordinate perceive legitimacy of the ruling power,[10] or hybrid.[11] An example of latter type and a point of reference for this paper is Beetham's conception, that acknowledges legality, i.e. the necessity of exercising power according to established rules, as an essential but insufficient aspect of legitimacy, contending that the power needs to be justified in terms of peoples' beliefs.[12]

10 Proceduralization impacts legitimacy in the following ways. Firstly, it advances the value of legality. Secondly, proceduralization legitimizes content moderation by promoting due process, an integral part of rule of law ideal,[13] that serves as a benchmark of political legitimacy[14] and an adequate framework for discussions about legitimate exercise of governance power.[15] Thus, imbuing the procedure with guarantees of due process enhances its legitimacy by aligning content moderation with rule of law.

C. Proceduralization by Platforms

11 In the area of copyright, proceduralization efforts of platforms appear to be most prominent in policymaking and oversight. Platforms devise increasingly detailed substantive and procedural rules on content moderation in terms of service, policies and help pages, approaching "the prolixity of a legal code".[16] While this approach might enhance legitimacy by offering users greater certainty, empirical evidence indicates that the proliferation of regulations has led to heightened complexity.[17] This is evident in the significant surge in the variety of documents, the gradual diversification in normative types and subjects of regulations.[18] Consequently, platforms achieve the opposite of the intended effect by making it challenging for users to navigate the waters of content moderation.

12 As to oversight, platforms have made efforts to facilitate external scrutiny of their content moderation practices through transparency reports. As an illustration, since December 2021, YouTube has been issuing a semi-annual Copyright Transparency Report.[19] These reports play a crucial role in promoting accountability and transparency by showcasing how content moderation decisions are made and enforced. Nevertheless, it's important to recognize that as platforms have the discretion

8 Douek (n 6). p. 3.

9 ibid.

10 Reinhard Bendix, *Max Weber: An Intellectual Portrait* (Routledge 1998).

11 Fabienne Peter, 'Political Legitimacy' (*Stanford Encyclopedia of Philosophy*, rev 2017 2010) <https://plato.stanford.edu/entries/legitimacy/#LegJusPolAut>.

12 David Beetham, *The Legitimation of Power* (Issues in Political Theory, Palgrave MacMillan 1991). p. 65-80. Beetham's framework includes the concept of subordinate consent as a component of legitimacy. However, it's important to note that this paper only partially employs his legitimacy framework as a reference point, and thus, the notion of subordinate consent is not a central focus within the scope of this paper.

13 Jeremy Waldron, 'The Concept and the Rule of Law' (2008) 43 Georgia Law Review <https://digitalcommons.law.uga.edu/cgi/viewcontent.cgi?article=1028&context=lectures_pre_arch_lectures_sibley>. p. 7, 62.

14 ibid. p. 3.

15 Nicolas Suzor, 'The Role of the Rule of Law in Virtual Communities' (2010) 25 Berkeley Technology Law Journal 1817. p. 1836.

16 Douek (n 6). p. 6.

17 João Pedro Quintais and others, 'Copyright Content Moderation in the EU: An Interdisciplinary Mapping Analysis' (reCreating Europe 2022) <http://dx.doi.org/10.2139/ssrn.4210278>.

18 ibid.

19 'Copyright Transparency Report H1 2021' (YouTube 2021) <https://blog.youtube/news-and-events/access-all-balanced-ecosystem-and-powerful-tools/>.

to determine which information is included and how it is presented, transparency reports can also be strategically leveraged to shape a specific narrative.[20] For example, platforms might use these reports to craft a more favourable image of their content moderation efforts.

13 Before the enactment of the pertinent EU legislation, namely the CDSM Directive, a significant proceduralization endeavour involved platforms voluntarily adhering to codes of conduct, which influenced creation and application of content moderation rules.[21] One specific aspect of codes of conduct that contributed to proceduralization were rules about notice and takedown mechanisms.[22] These processes were notably absent from the EU safe harbour framework at that time.[23] Furthermore, codes of conduct encompassed obligations such as issuing warnings to subscribers engaged in infringing activities, retaining crucial traffic data, disclosing the identities of implicated subscribers and terminating accounts of the infringers.[24]

14 Another example of voluntary proceduralization initiative are Santa Clara Principles, are a set of guidelines developed to safeguard freedom of expression and privacy rights in content moderation practices and endorsed by major platform providers such as Meta, Google, Reddit, X, and GitHub.[25] The principles emphasize transparency, accountability, and user empowerment in online platforms' content removal policies and advocate for clear explanations of content moderation decisions, opportunities for appeal, and limitations on the use of automated tools in content removal.

15 Providing a possibility to appeal platform's content moderation actions may also be counted among the proceduralization measures. The problem is that platforms partially do so to comply with their legal obligations, in particular DMCA.[26] However, it's worth noting that many platforms proactively take the initiative to establish complaint and redress mechanisms that go beyond what is strictly required by law.[27] To that extent, provision of such mechanisms may be considered platforms' own proceduralization initiative.

16 The crown jewel of platforms' proceduralization efforts is Meta's Oversight Board, that gives impression of being created for the sole purpose of legitimation. Its design borrows attributes of supreme or constitutional courts,[28] creating "an institutional aesthetic of governance."[29] Oversight

20 Aleksandra Urman and Mykola Makhortykh, 'How Transparent Are Transparency Reports? Comparative Analysis of Transparency Reporting across Online Platforms' (2023) 47 Telecommunications Policy 102477.

21 For instance, in 2007, several UGC platforms, such as MySpace, Veoh, DailyMotion, and Soapbox, joined forces with major players of creative industry such as Disney, CBS, NBC Universal, and Viacom to put forth a set of guidelines known as the 'Principles for User Generated Content Services' See 'User Generated Content Principles' <http://ugcprinciples.com/> accessed 7 June 2023; discussed in Michael S Sawyer, 'Filters, Fair Use & Feedback: User Generated Content Principles and the DMCA' (2009) 24 Berkeley Technology Law Journal 363.

22 P Bernt Hugenholtz, 'Codes of Conduct and Copyright Enforcement in Cyberspace' (7 March 2012) <https://papers.ssrn.com/abstract=2017581> accessed 7 June 2023.

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

24 Hugenholtz (n 22).

25 'Santa Clara Principles on Transparency and Accountability in Content Moderation' (*Santa Clara Principles*) <https://santaclaraprinciples.org/images/santa-clara-OG.png> accessed 6 March 2024.

26 Digital Millennium Copyright Act (DMCA), 17 U.S.C. § 1201 et seq. (1998).

27 Péter Mezei and István Harkai, 'End-User Flexibilities in Digital Copyright Law – An Empirical Analysis of End-User License Agreements' (3 July 2021) <https://papers.ssrn.com/abstract=3879740> accessed 14 September 2023.

28 For example, the case selection mechanism bears resemblance to the certiorari process employed by the US Supreme Court. This process involves the careful selection of a limited number of cases, with a particular emphasis on disputes that present significant legal questions. Inspiration from European constitutional courts is on the contrary visible in "a prevalence of written over oral submission, a limited role for the disputing parties, and an emphasis on the development of the law for the future. See Goanta and Ortolani (n 7).but fail to ensure adequate access to justice through content moderation when harms arise. This chapter focuses on a gap in current scholarship on platform governance, by addressing content moderation from the procedural perspective of dispute resolution. We trace the emergence of content moderation as a form of digital dispute resolution, proposing a theoretical framework for the understanding of social media platforms as private adjudicators, and illustrating how platforms have progressively embraced this role. This framework is further complemented by an empirical overview of the content reporting mechanisms of four social media platforms (Facebook, TikTok, Twitch and Twitter p. 17.

29 Monroe E Price and Joshua M Price, 'Building Legitimacy

Board serves as policy advisor, an appeal board and a source of information about Meta's content moderation structures and processes. It remedies ad hoc content moderation shortcomings by reviewing a small number of "highly emblematic" cases selected by it from appeals by users,[30] assessing the compliance of content with Facebook's policies[31] in the light of international human rights standards.[32] Oversight Board has yet to make a decision on any copyright-related matter. However, the possibility remains that it may do so in the future, for instance in a case involving thebalancing of copyright and freedom of expression.

D. Proceduralization in Regulatory Framework

I. Article 17 CDSM Directive

17 The first major regulatory intervention of the EU legislator concerning copyright content moderation is Article 17 of the CDSM Directive.[33] Procedural elements in Article 17 give the impression of being somewhat perfunctory. Article 17 enhances due process for rightsholders by providing them with an additional avenue of asserting their rights by means of providing "relevant and necessary" information regarding their works.

18 However, Article 17 does not improve the position of the users from the procedural perspective. An interesting safeguard is the obligation of platforms to inform their users in their terms and conditions of the possibility to use the defence of copyright exceptions or limitations. Unfortunately, it seems to be dysfunctional, as explained in Section 5.1.2.

19 Mechanisms for appealing content moderation decisions — a single procedural safeguard of relevance for users — are lacking. Firstly, the range of content moderation decisions which may be appealed is limited to removal or access restriction, not taking into account that the preferred action in the majority of copyright infringement claims is demonetization.[34] Secondly, Article 17 (9) does not provide foundations for adversarial proceedings.[35] For example, it merely requires that decisions to disable access to or remove uploaded content shall be subject to human review, without specifying who performs the review. Consequently, complaint and redress mechanisms offered by Meta (for Facebook and Instagram) and by YouTube,[36] in which the platform acts as a messenger rather than an arbiter and the decision about content is made by the rightsholder,[37] would be compliant with this provision. Another remedy available to the user, out-of-court redress mechanisms, should enable impartial settlement of disputes arising from content moderation. Article 17 places no requirements on the dispute resolution bodies and does not incentivize either platforms or rightsholders to participate in the scheme.

20 Additionally, Article 17 does not promote the quality of content moderation decisions. It does not attempt to influence the accuracy of content moderation decisions[38] by placing requirements on the setting

in the Absence of the State: Reflections on the Facebook Oversight Board' [2023] International Journal of Communication; Vol 17 (2023) 3 p. 6.

30 'Oversight Board | Independent Judgment. Transparency. Legitimacy.' <https://www.oversightboard.com/> accessed 8 May 2023.

31 ibid.

32 Those norms include the International Covenant on Civil and Political Rights (ICCPR)'s Article 19, which states that while "everyone shall have the right to freedom of expression...the exercise of [that] right may...be subject to certain restrictions, but only...as provided by law and are necessary. 'Oversight Board Annual Report 2021' (Oversight Board 2022). p. 9.

33 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

34 Henning Grosse Ruse-Khan, 'Automated Copyright Enforcement Online: From Blocking to Monetization of User-Generated Content', *Transition and Coherence in Intellectual Property Law: Essays in Honour of Annette Kur* (Cambridge University Press 2021) <10.1017/9781108688529>. p. 2.

35 A cornerstone to the right to a fair trial, a corollary to the right to an effective remedy according to Article 47 of the Charter of Fundamental Rights. Manuel Kellerbauer, Marcus Klamert and Jonathan Tomkin (eds), *The EU Treaties and the Charter of Fundamental Rights: A Commentary* (Oxford University Press 2019) <https://doi.org/10.1093/oso/9780198794561.001.0001> accessed 16 April 2023. p. 2222.

36 'Dispute a Content ID Claim - YouTube Help' <https://support.google.com/youtube/answer/2797454?hl=en-GB> accessed 14 April 2023.

37 'Resolve Usage Disputes in Rights Manager' (*Meta Business Help Centre*) <https://en-gb.facebook.com/business/help/2523148971045474> accessed 14 September 2023.

38 Niva Elkin-Koren, 'Fair Use by Design' (2017) <https://papers.ssrn.com/abstract=3217839> accessed 6 January 2023.

of parameters of automated content filtering tools, nor does it require the original content moderation decision or the result of dispute to be accompanied by justification. Transparency is also neglected by Article 17. The only transparency obligation in Article 17 does not extend to dispute resolution. It is limited to information regarding platforms actions towards content and the use of licensed works in content. Additionally, it is curiously one-sided, applying only to the rightsholder. Therefore, the regulator is unable to supervise the quality of content moderation due to the lack of data.

II. Digital Services Act

21 On the contrary, the DSA, which marks a "procedural turn" in EU lawmaking, considerably proceduralizes content moderation by introducing a set of obligations spanning the whole content moderation process.[39] Article 16 DSA establishes clear rules for reporting content. Article 17 DSA, a provision that is also applicable to the Article 17 CDSM regime, requires every content moderation decision to be accompanied by a statement of reasons. It should explain what actions are being taken and their scope, as well as where and for how long they apply, the reasons for the decision, use of automated processes and legal basis for determining that the piece of content in question is illegal. Importantly, it should also contain information about how the recipient of the decision may seek redress.

22 Article 20 DSA broadens access to justice by encompassing a significantly wider array of content moderation decisions that extend beyond the mere blocking and removal of content.[40] It emphasizes accessibility and fairness: submission of complaints should occur electronically and free of charge, the mechanism should be user-friendly and complaints should be handled in a non-discriminatory, diligent and non-arbitrary manner. Also, according to Article 14 DSA, the platform should provide rules of the complaint-handling procedure in Terms and Conditions.

23 The out-of-court dispute settlement mechanism in DSA is also fully compliant with the proceduralization approach. To fall within the purview of Article 21, a dispute resolution body must obtain certification, contingent on criteria such as impartiality and independence, the establishment of clear and fair procedural rules, and the capacity to efficiently resolve disputes[41] – all of which align with the due process requisites specified in Article 47 of the EU Charter of Fundamental Rights. Furthermore, the designated body must possess the requisite expertise, and the dispute should take place online. [42] The platform is mandated to engage in the dispute resolution process presented by an entity chosen by the service recipient, unless a dispute has already been resolved concerning the same information and the same grounds. [43] Moreover, the DSA imposes a time constraint of 90 days for the resolution process[44] and establishes a mechanism for attributing procedural costs, which tilts the balance in favour of users over platforms,[45] contributing to equality of arms.

24 The same holds for transparency provisions regarding use of automated content recognition tools,[46] complaint and redress mechanisms,[47] cases submitted to out-of-court dispute resolution bodies[48] and database of content moderation decisions.[49] They provide the public with exhaustive information

39 Pietro Ortolani, 'If You Build It, They Will Come: The DSA "Procedure Before Substance" Approach', *Putting the Digital Services Act into Practice: Enforcement, Access to Justice, and Global Implications* (Verfassungsblog 2023).

40 According to Article 20(1) DSA, users can appeal decisions regarding the removal or restriction of access to content, the suspension or termination of services or user accounts, and the restriction of monetization of user content.

41 Article 21(3) DSA.

42 ibid.

43 Article 21(2) DSA.

44 Article 21(2) DSA.

45 According to Article 21(5) DSA, if the out-of-court dispute settlement body decides in favor of the user, the online platform provider must bear all fees and reimburse the user for reasonable expenses related to the dispute. If the decision favors the provider, the user is not required to reimburse any fees or expenses of the provider of the online platform paid, unless they are found to have acted in bad faith.

46 These include the following obligations of providers of online platforms: include information about the use of algorithmic decision-making in content moderation in their Terms and Conditions (Article 14(1) DSA); provide detailed information about use of automated tools in content moderation in the annual transparency report (Article 15(1)(c) and (e) DSA); and inform a user in the particular instance of content moderation about the use of the use made of automated means in taking the decision regarding content (Article 17(3) c) DSA).

47 Article 20 DSA.

48 Article 21 DSA.

49 Article 24(5) DSA.

regarding both types of mechanisms, enabling the exercise of control and promoting consistency of decision-making.

III. Limits of Proceduralization Approach

25 As previously discussed, proceduralization significantly enhances the legitimacy of content moderation practices. Additionally, it establishes legal certainty by defining expectations for all involved parties. The implementation of formalized procedures also simplifies the task of holding platforms accountable for their content moderation decisions, as these procedures are documented and subject to review and assessment. Nonetheless, it's important to acknowledge that proceduralization does have its limits.

26 As was said in Section 2, proceduralization promotes legitimacy of content moderation by advancing the legality principle. It is also important to bear in mind that legality is only one aspect of the legitimacy concept — necessary, but insufficient.[50] Relying solely on a formalistic approach cannot inherently legitimize content moderation. The reason is that *"[a]uthority also needs to be sociologically and morally legitimate to be accepted, and legalistic legitimacy alone is not enough to garner social and moral respect".*[51] Content moderation should also be justifiable in terms of beliefs of the relevant constituency,[52] who in this case arguably are the users as the addressees of platform governance.

27 The second problem is that the impact of mechanisms of redress, which constitute an essential vehicle of proceduralization approach, is dependent on the uptake by the stakeholders – civic rights organisations, external dispute settlement bodies, but most importantly ordinary users. For both of those mechanisms, the uptake by individuals is crucial, that happens to be notoriously low in copyright content moderation cases.[53] Possible reasons for under-assertion include intimidation and a weak prospect of successful redress.[54] A causality circle emerges here: the mechanisms will serve as an instrument of legitimation when taken up by the people, and the people, in turn, will adopt these mechanisms if they perceive them as a legitimate means of resolving their problems.

E. Procedural Justice Approach as a Successor of Proceduralization

28 To address the concerns with lukewarm adoption of dispute resolution mechanisms and enhance the legitimacy of content moderation in the sociological sense, this section proposes a complementary legitimization strategy that aligns with the proceduralization approach. The strategy entails the incorporation of the psychological concept of procedural justice into the practical implementation of the legal framework, specifically focusing on the establishment of dispute resolution mechanisms.

I. Procedural Justice, Due Process and Legitimacy

29 Procedural justice in the psychological sense refers to how individuals subjectively perceive the fairness of the process. While distinct from distributive justice that centres on outcome fairness, procedural justice nonetheless is empirically proven to have a positive impact on distributive justice judgments, even in cases when outcomes are negative.[55] The scope of procedural justice concept is very broad – in fact, it is applicable to any social processes where outcomes are allocated,[56] which distinguishes it from formal due process principles and makes it suitable for application to content moderation. At the same time, procedural justice is a natural pendant to due process principles. The popular notion of fair procedure provided the original impetus for the creation of due process principles, while due process principles in turn equip people with "a helpful template for what fair process looks like" in forming the perception of what is fair.[57]

50 Beetham (n 12). p. 69.

51 Douek (n 6). p. 15.

52 Beetham (n 12). p. 17.

53 Annemarie Bridy and Daphne Keller, 'U.S. Copyright Office Section 512 Study: Comments in Response to Notice of Inquiry' (2017) 7 SSRN Electronic Journal; Urban, Karaganis and Schofield (n 5).

54 Fiala and Husovec (n 5).

55 Edgar Allan Lind and Tom Tyler, *The Social Psychology of Procedural Justice*, vol 18 (Springer Science + Business Media LLC 1988). p. 67.

56 Rebecca Hollander-Blumoff and Tom Tyler, 'Procedural Justice and the Rule of Law: Fostering Legitimacy in Alternative Dispute Resolution' (2011) 2011 Journal of Dispute Resolution <https://scholarship.law.missouri.edu/jdr/vol2011/iss1/2/>.

57 ibid. p. 9.

30 Various criteria influence procedural justice judgments. For instance, it has been found that people value control over the process and outcome, ethical behaviour of the authority, and impartiality.[58] Ethicality encompasses politeness and respect for disputants' rights, while process control involves being heard and presenting information that the individual considers important. Other sources cite the authority's consideration of arguments,[59] ability to gather sufficient information for a high-quality decision, consistency in decisions, and credibility of the decision-making authority in the sense that it made best efforts to do the disputants justice.[60] Additional criteria which matter to the disputants include airing the problem, speed of resolution, personal control, animosity reduction, cost, minimizing disruption of everyday affairs, and reducing the possibility of future conflict.[61]

31 The connection between legitimacy and procedural justice is supported by empirical evidence showing that people base their judgments about the overall legitimacy of authorities on their personal experiences with their representatives.[62] While various factors impact people's attitudes towards authorities,[63] assessments of procedural fairness have been identified as the major influence,[64] surpassing distributive fairness.[65] Notably, even in cases of negative outcomes, fair procedures act as a cushion, maintaining high levels of support for the authority.[66]

II. Procedural Justice Values in Content Moderation

As regards empirical evidence of which elements of procedural justice are relevant for content moderation, it is possible to draw from a rich body of knowledge has emerged in recent years through empirical studies examining user accounts of their interactions with platforms.[67] Since the literature focuses on shortcomings of content moderation, these accounts serve to define the procedural justice values in content moderation negatively, i.e. by their absence.

Value	Content Moderation Stage
"Legal aid" – explanation of substantive and procedural platform policies, ideally with examples	Stage 1
Individualized explanation of the decision	Stage 1
Accessibility of redress mechanisms	Between Stage 1 and 2
Quality of human interactions	Stages 1 and 2
Opportunity to present user's case	Stage 2
Impartiality of content moderators	Stage 2
Qualification of content moderators	Stage 2

This chart summarizes procedural justice values derived from the studies, explained in more detail in the following sections.

1. The First Stage – Content Detection

32 The bulk of academic literature focuses on the first content moderation stage – (automated) detection of content that infringes copyright or platform

58 Tom R Tyler, 'What Is Procedural Justice?: Criteria Used by Citizens to Assess the Fairness of Legal Procedures' (1988) 22 Law & Society Review 103.

59 Donald E Conlon, E Allan Lind and Robin I Lissak, 'Nonlinear and Nonmonotonic Effects of Outcome on Procedural and Distributive Fairness Judgments.' (1989) 19 Journal of Applied Social Psychology 1085.using a classic procedural justice paradigm (e.g., L. Walker et al; see record 1975-23047-001

60 Hollander-Blumoff and Tyler (n 56). p. 5.

61 Robin I Lissak and Blair H Sheppard, 'Beyond Fairness: The Criterion Problem in Research on Dispute Intervention.' (1983) 13 Journal of Applied Social Psychology 45.

62 Tom R Tyler, *Why People Obey the Law* (Princeton University Press 2006) <http://www.jstor.org/stable/j.ctv1j66769> accessed 5 October 2022. p. 88, 91.

63 For example, previous experience, social background, moral convictions of the individual and instrumental considerations, such as personal gain from the outcome.

64 Lind and Tyler (n 55). p. 78.

65 ibid. p. 65.

66 ibid. p. 71.

67 Sophie Bishop, 'Managing Visibility on YouTube through Algorithmic Gossip' (2019) 21 New Media & Society 2589; Sophie Bishop, 'Influencer Creep: How Artists Strategically Navigate the Platformisation of Art Worlds' [2023] New Media & Society 14614448231206090; Laura Savolainen and Minna Ruckenstein, 'Dimensions of Autonomy in Human–Algorithm Relations' [2022] New Media & Society 14614448221100802; Sarah Myers West, 'Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms' (2018) 20 New Media & Society 4366; Kelley Cotter, 'Playing the Visibility Game: How Digital Influencers and Algorithms Negotiate Influence on Instagram' (2019) 21 New Media & Society 895; Brooke Erin Duffy and Colten Meisner, 'Platform Governance at the Margins: Social Media Creators' Experiences with Algorithmic (in)Visibility' (2023) 45 Media, Culture & Society 285.

policies and application of a wide range of content moderation measures, including restriction of visibility or demotion, which forms the primary object of interest in the studies. Many of them describe how users, in particular content creators, attempt to decode and adapt to the principles of functioning of algorithms and avoid having the visibility of their content reduced, [68] while some of them examine how algorithms shape the creative process and the presentation of users on the internet.[69]

33 The first element that emerges from the user accounts is the need for clear and detailed rules of application of platform "substantive law". Users perceive rules contained in terms and conditions or community guidelines as vague and unhelpful and miss specific examples.[70] Therefore, they develop heuristics, such as which hashtags to use or how much skin to show to avoid being flagged for nudity, and share this information in support groups.[71] In the field of copyright, users have proven themselves woefully ignorant of the legal basics and platform policies, expressing a desire to learn more.[72] Consistency in platform decisions, unsurprisingly, emerges as another trait valued by users, who frequently expressed frustration at the erratic nature of platform decisions.[73]

34 Another important factor is an individualized explanation of reasons behind the decision. Users lamented the lack of detailed explanation of how user violated community guidelines, reporting that instead, they receive generic repetitive references to general platform policies.[74] Unfortunately, the obligation to provide statement of reasons for the decision introduced by Article 17 DSA is unlikely to change the users' dissatisfaction in that regard, since the provision merely lists the mandatory elements without requesting an individualized response. As the examples from DSA Transparency Database demonstrate, platforms continue to use formulaic responses.[75]

68 Cotter (n 67); Duffy and Meisner (n 67).

69 Cotter (n 67).

70 Duffy and Meisner (n 67). p. 295.

71 ibid. p. 297.

72 Daria Dergacheva and Christian Katzenbach, '"We Learn Through Mistakes": Perspectives of Social Media Creators on Copyright Moderation in the European Union' (2023) 9 Social Media + Society 20563051231220329. p. 5.

73 Duffy and Meisner (n 67).

74 ibid.

75 https://transparency.dsa.ec.europa.eu/statement

2. Stage One-and-a-Half: Transition from the First to the Second Stage

35 For the success of next content moderation stage, the internal and external redress mechanisms, the decisive moment is whether users will engage with them. Therefore, accessibility emerges as a prerequisite value for these mechanisms. This is corroborated by the evidence from content moderation, citing that a relatively high number of users express desire to appeal the mechanism and yet encounter problems such as unclear instructions,[76] and an example from a different field – the soon-to-be repealed ODR platform for resolution of consumer disputes, which, while exhibiting a 8.5 million visits, only enables on average 200 cases per year to be treated by ADR entities,[77] since its design is confusing to users.[78] While DSA attempts to address this problem by requiring that the user accesses the procedure simply by clicking on a link that leads to internal mechanism or a page where dispute settlement bodies present themselves for an easy selection.[79]

3. The Second Stage – Appeal Mechanisms

36 The second stage becomes relevant when the content is blocked and the user appeals the decision. Both Article 17 of the DSM Directive and Article 20 of the Digital Services Act provide and obligation of platforms to establish an internal complaint-handling mechanism, where the platform acts as an arbiter and, when platform's own-initiative content moderation measure is disputed, platform plays the party to the dispute. The use of external mechanism is not pre-conditioned on the internal process.

37 An overarching and essential factor for the second content moderation stage is human interaction. This factor related to both the desire to be heard, i.e. to present information the individual considers important, and to receive a satisfactory explanation of their case. Some users went to considerable lengths to exercise their "right to be heard" - finding other means of communication not designed for such cases, such as via company accounts on other social media platforms or technical support channels.[80] Nevertheless, the users were not willing to accept just any human interaction; it had to meet specific quality standards. Some users who interacted with human personnel complained that their responses were formulaic and repetitive, not offering any relief in comparison with responses form a bot.[81] Another concern was over the qualification and impartiality of content moderators. The users expressed doubts about content moderators' expertise and impartiality, asserting that they are biased towards marginalized groups.[82]

III. The Role of the Regulator

38 It remains to be examined how can the regulator contribute to introducing procedural justice in the design of redress mechanisms, using the above-described procedural justice index. In case of out-of-court dispute resolution bodies, Digital Services Coordinators ("DSC")[83] have a considerable leverage over them, since they are the authority which provides them with time-limited and revocable certification, assessing inter alia whether their rules of procedure are fair or whether the body's expertise allows them to settle the dispute effectively.[84] Further, the bodies report to DSC annually as regards their operation and DSC may offer them recommendations as to how improve their functioning.[85] In both of these

76 Myers West (n 67). p. 4378.

77 Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL repealing Regulation (EU) No 524/2013 and amending Regulations (EU) 2017/2394 and (EU) 2018/1724 with regards to the discontinuation of the European ODR Platform 2023.

78 Emma van Gelder, *Consumer Online Dispute Resolution Pathways in Europe: An Analysis into Standards for Access and Procedural Justice in Online Dispute Resolution Procedures* (2022) p. 158-161.

79 This is required by the DSA in several provisions. Firstly, article 17(3)f) requires statement of reasons to contain "clear and user-friendly information on the possibilities for redress available to the recipient of the service in respect of the decision, in particular, where applicable through internal complaint-handling mechanisms, out-of-court dispute settlement and judicial redress". Secondly, article 20(3) demands that the internal complaint-handling mechanism is easy to access and user-friendly. Also, article 21(1) requires providers of online platforms to ensure that "information about the possibility for recipients of the service to have access to an out-of-court dispute settlement, [...], is easily accessible on their online interface, clear and user-friendly".

80 Myers West (n 67). p. 4376.

81 ibid. p. 4377.

82 Duffy and Meisner (n 67). p. 238.

83 The authorities responsible for enforcement of DSA, together with the Commission. See Articles 49-51 DSA.

84 Article 21(3) DSA.

85 Article 21(4) DSA.

functions, DSC may use the procedural justice index as a point of reference.

39 The key question is how to encourage platforms to prioritize procedural justice when shaping their dispute resolution mechanisms. While platforms are showing engagement with proceduralization trend, it remains uncertain whether their commitment is sincere or a form of virtue signalling. Given their profit-oriented nature, platforms might concentrate on improving content moderation in less controversial areas than copyright, where the discourse is dominated by two antagonist groups of rightsholders and free speech advocates.

40 In case of very large online platforms, Commission and the European Board for Digital Services[86] may impact their implementation of relevant DSA provisions by influencing the standards for adequate risk mitigation measures based on the above index. As was mentioned above, very large online platforms are under obligation to mitigate systemic risks stemming from the design or functioning of their service and its related systems.[87] Such risks include "any actual or foreseeable negative effects for the exercise of fundamental rights,"[88] which covers the right to an effective remedy and to a fair trial. Content moderation features both as a factor to be taken into consideration in risk assessment[89] and as the object of risk mitigation measures.[90]

41 Commission may provide guidelines on measures relating to specific risks[91] and adopt delegated acts laying down the necessary rules for the performance of the annual audits by independent organisations, which assess among other things compliance with due diligence obligations, including operation of the internal redress mechanism.[92] The Board is expected to identify best practices for risk mitigation in its yearly reports.[93] Ideally, the concerted efforts of Commission, DSC and the Board should be directed toward creating an index of parameters that will be used to assess the adequacy of the mechanisms.

F. Conclusion

42 To summarize the above findings, although proceduralization as a legitimizing strategy in platform governance has its merits, it addresses only one facet of legitimacy — legality, neglecting legitimacy in the sociological sense. This deficit can be mitigated by a complementary legitimation strategy, namely through incorporating empirically derived values of procedural justice into the mechanisms mandated by the CDSM and DSA. To facilitate this integration, an index outlining procedural justice values pertinent to users should be developed. While this paper has provided a preliminary framework of such values within the context of content moderation, further research is warranted, as these values were derived from studies with slightly different objectives.

43 In conclusion, the paper has provided an analysis of content moderation proceduralization and outlined potential future directions. The hope is that this exploration contributes to the ongoing discourse on the regulation of online platforms and the advancement of effective governance strategies.

86 An independent advisory group of Digital Services Coordinators on the supervision of providers of intermediary services. Its tasks are contributing to the consistent application of DSA, coordinating and contributing to guidelines and analysis of the Commission and Digital Services Coordinators and other competent authorities and assisting the Digital Services Coordinators and the Commission in the supervision of very large online platforms. See Article 61 DSA.

87 Articles 34 and 35 DSA.

88 Article 34(1)(b) DSA.

89 Article 34(2)(b) DSA.

90 Article 35(1)(c) DSA.

91 Article 35(3) DSA.

92 Article 37(7) DSA.

93 Article 35(2)(b) DSA.