HISTORIES AND RUPTURES IN PLATFORM GOVERNANCE

Tomás Guarna
Stanford University

Emillie de Keulenaar
University of Copenhagen

Anna Gibson
Massachusetts Institute of Technology

Diyi Liu
University of Copenhagen

Scholars of platform governance, referring to dynamic relationships between governments, platform companies, and civil society (Gorwa 2019) have often inquired the *new ways* that online activity is structured by emerging technologies, norms, and political tensions (e.g. Gorwa 2024; Suzor 2020; Van Dijck 2021; Crawford and Gillespie 2016; Bloch-Wehba 2019; Caplan and Gillespie 2020). The emphasis on novelty is certainly warranted, as platformization continuously introduces new areas for political contestation. At the same time, we recognize a need to historicize these power relations: tracing the continuities, reconfigurations, and above all *ruptures* among novelties. To achieve this, we put in conversation four papers that advance our historicization of platform governance.

Core to our intention is the notion that speech moderation "materializes everywhere" as a practice that constitutes and defines the boundaries of public spheres (Robert C. Post, 1998: 2). Societies that opt not to rely solely on regulatory means, such as state censorship, seek to maintain or restore public order by modulating the circulation of "excesses" relative to a historical norm (Pohjonen and Udupa 2018: 1174). Examples include hate speech, memory and blasphemy laws, interpersonal speech norms or more complex historical processes (denazification, decommunization, decolonisation), which, amongst other objectives, seek to drive historical change by re-educating, marginalising or eradicating the language through which undesired behaviour, ideas or beliefs — past or present — may subsist. Such processes are not always formalized by law (Faugere et al., 2023) but permeate in public norms and inform platform policies and techniques.

What is considered acceptable speech has of course changed dramatically across various public debate, such that speech norms are "in constant negotiation" (Fraser, 1990) — for example, over what aspects of history should or should not be allowed to be conjured by language (see, e.g., Vincent (2008)). Across these media environments, the maintenance of consensus in such norms is not always guaranteed, leading at times to profound *ruptures* both *within* and *across* media environments. In the latter case, the absence of strong consensus-building affordances in online platforms has seen profound ruptures into alternative platforms with competing moderation philosophies. There have been profound ruptures within single platforms such as X or Meta in the former. These ruptures result partly from deeply contentious public debates about the very *norms* of public discussion, which modulate the boundaries of social media platforms as public spheres*.*

The four papers in this panel examine historicise ruptures in platform governance. Two explore "prehistories": one tracing content moderation's emergence at eBay through leadership transitions and another comparing CB radio with CompuServe's CB Simulator to demonstrate early digital content moderation. These papers illuminate how content moderation responded to commercial and political pressures, marking a shift from idealistic visions of unregulated online spaces. The other two papers offer recent analyses of contemporary ruptures. One examines how media governance approaches historically shape platform regulation in morally conservative contexts, revealing tensions between democratic aspirations, state control, religious considerations, and commercial interests. The final paper investigates the shift in content moderation from simple adjudication to complex forms of mediation and consensus-building. Putting these latter two papers in conversation allows us to trace the evolving nature of platform governance across different cultural contexts and moderation approaches.

All four papers, drawing from diverse historical and geographical contexts, offer glimpses into distinct ruptures: pivotal moments when the collective understanding of "acceptable speech" underwent imperative reconsideration, becoming inscribed (see Akrich 2010) through the affordances of evolving digital platforms. By juxtaposing prehistories with more contemporary analyses, we can effectively historicize the ongoing evolution of platform governance, shedding light on how challenges, tensions, and "fixes" have diminished and persisted over time. Examining the historical contingency of these "ruptures" also invites us to consider the future of platform governance. As civil society initiatives begin to conceptualize novel approaches to platform organization (see Schneider 2024; Zuckerman 2020) or even the very purpose of moderation (Tang, 2024), recognizing the historical context encourages reflection on potential "otherwises" in platform governance.

References

Akrich, M. (2010). The de-scription of technical objects. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (Nachdr.). MIT Press.

Caplan, R., & Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media + Society, 6*(2), 205630512093663. https://doi.org/10.1177/2056305120936636

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society, 18*(3), 410–428. https://doi.org/10.1177/1461444814543163

Gorwa, R. (2024). *The politics of platform regulation: How governments shape online content moderation* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780197692851.001.0001

Schneider, N. (2024). *Governable spaces: Democratic design for online life* (1st ed.). University of California Press.

Suzor, N. (2020). A constitutional moment: How we might reimagine platform governance. *Computer Law & Security Review, 36*, 105381. https://doi.org/10.1016/j.clsr.2019.105381

Tang, A. (2024, November 26). Embrace the shift to 'prosocial media.' *Wired.* https://www.wired.com/story/prosocial-media-social-networks-discourse-decentralization

Van Dijck, J. (2021). Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society, 23*(9), 2801–2819. https://doi.org/10.1177/1461444820940293

Zuckerman, E. (2020). The case for digital public infrastructure. *Knight First Amendment Institute.*

FROM THE GIFT ECONOMY TO "TRUST AND SAFETY": EBAY (1995-2004) AND THE RISE OF PLATFORM GOVERNANCE

Tomás Guarna
Stanford University

In July 2004, Memphis judge and founder of the National Civil Rights Museum D'Army Bailey tried to auction the bathtub where Martin Luther King Jr.'s assassin had stood on eBay, by then the world's most popular online marketplace. eBay's "Trust and Safety" department, the company's arm in charge of removing content that infringed the site's policies, removed the listing, citing a policy against items "that graphically [portrayed] violence or victims of violence, and [lacked] substantial social, artistic or political values" (Baird, 2004). Bailey, who had been aiming to raise funds for the Museum, asked the company for the listing to be brought back. eBay's Trust and Safety team now faced a difficult question for the application of their policies: Should determinations be based

solely on the contents of the listing, or should they also consider the identity of the user making the listing? The case reached the team's "gray items group," a small committee of employees that analyzed unclear cases. One employee argued that eBay was not in the business of making judgments of people, but of items, and that this could snowball into requiring complex evaluations for much more policy infringement cases (eBay employee, 2023, personal communication). After hours of deliberation, the committee decided that Bailey's listing "historic value," and that therefore it qualified for a caveat that made it admissible under eBay's policies (Baird, 2004).

This episode shows how, by 2004, eBay employed a sophisticated system – including norms, processes, and expertise – to make decisions over what content was admissible on its website. It is also emblematic of a radical shift on how content governance was understood by the company's leadership. When Pierre Omidyar started the company in 1995, he modeled the site on early virtual communities like Usenet, where allegedly implicit norms of reciprocity governed an online "gift economy," referring to anthropologist Marcel Mauss' (2002 [1925]) concept of non-monetary exchange systems. This was anchored in a deep techno-utopianism (see Turner 2006). When Meg Whitman took reins of the company in 1998, this hands-off vision largely disappeared. As the bathtub case demonstrates, during Whitman's tenure, eBay developed a technical and organizational machinery to regulate and surveil user activity.

This machinery is an early case of the industrial practice contemporarily known as "commercial content moderation," which information science scholar Sarah Roberts (2022) defines as "the organized practice of screening user-generated content posted to internet sites, social media, and other online outlets" (33). Contemporary commercial content moderation involves formalized policies, which stipulate what users can and cannot do, and mechanisms of enforcement, involving manual labor and algorithmic systems (see Roberts 2019; Gorwa et al., 2020). As communication scholar Tarleton Gillespie (2018) argues, it is a resource-intensive practice that involves inescapable judgements about values, as companies must constantly decide what constitutes acceptable speech and harmful content. While content removals often face public backlash as censorship, platforms that host user content must continue moderating, as they see content curation as essential to their business.

This paper provides a glimpse at the development of content moderation at eBay. Through analyzing news coverage, external and internal company documents, and popular press books (ex. Cohen 2022; Whitman and Hamilton 2010) about the company, it examines how eBay developed a set of practices and expertise around content moderation, which would come to be named "Trust and Safety." Trust and Safety was both a practice and a theory, as it came to frame an elaborate philosophy that positioned the company as a guarantor of free commerce and as a responsible societal actor. The paper traces how Trust and Safety developed organizational changes within eBay, shedding light on how American business culture shaped content governance.

eBay's story is useful as a case to understand how platform governance rose alongside the commercialization of the Internet (Greenstein 2015). The rise of content moderation is often traced to an increase in the scale of users in online services. Communication

scholar Jessa Lingel (2020) calls the development of formal rules in online spaces the "gentrification" of the Internet, which happened because "more people came online and new platforms sprouted to meet their needs" (2). In a similar vein, in his analysis of the emergence of formal rules in eBay, economic sociologist Vili Lehdonvirta (2022) argues that the company developed processes to regulate user activity as the marketplace "kept growing and changing" (49), and new challenges (for example, the sale of deadly drugs) emerged. Therefore, the company developed rules and policies not "because eBay's executives were hungry for power," but "only begrudgingly, in response to market failures that could not be otherwise addressed" (52). This paper shows how this narrative provides only a partial view of the development of content governance. Indeed, it traces how eBay did not develop its content moderation machinery only in response to user activity, but to a broader ecosystem of stakeholders, including regulators, other businesses, and the company's board of directors.

By following Trust and Safety's development alongside eBay's transformation from a start-up to a multinational corporation, this paper shows how content governance emerged inseparably with the novel pressures of the commercialized Internet. I start by tracing how the privatization of the internet led to the emergence of commercial online services, which enabled the governance at the level of content (Zittrain 1997; Galloway 2006). Then I move to historically contextualizing the emergence of content moderation at eBay, examining the transition from Pierre Omidyar's vision of community self-governance to Meg Whitman's implementation of centralized content moderation. By examining a unique archival material, the *eBay Trust Playbook*, alongside developments within the company, the following sections develop two key insights. First, eBay developed Trust and Safety as a body of knowledge, practices, and expertise. Second, this philosophy understood "trust" not just as an engineerable relation between users, but also as an ethical relationship between the company and a wider network of stakeholders. This understanding of trust explains the emergence of content policies. Finally, I connect eBay's development of Trust and Safety with the discursive work of web companies to portray their sites as "platforms," which, according to Tarleton Gillespie (2010) allows them to justify their services as neutral conduits for user expression while exerting control and curatorship. The episodes at eBay show how Trust and Safety legitimized the company to portray itself as neutral, while remaining compatible with the moral politics of American business culture.

References

Baird, W. (2004, July 16). eBay removes bathtub used in King assassination from its internet listings. *Associated Press International.* NexusUni. https://advance.lexis.com/api/document?collection=news&id=urn%3acontentItem%3a4CW7-PY60-00BT-M046-00000-00&context=1519360&identityprofileid=7N5PFF56124

Cohen, A. (2002). *The perfect store: Inside eBay.* Little, Brown and Co.

eBay employee. (2023). Personal communication.

Galloway, A. R. (2006). *Protocol: How control exists after decentralization.* MIT Press.

Gillespie, T. (2010). The politics of "platforms." *New Media & Society, 12*(3), 347–364. https://doi.org/10.1177/1461444809342738

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1), 2053951719897945. https://doi.org/10.1177/2053951719897945

Greenstein, S. M. (2015). *How the internet became commercial: Innovation, privatization, and the birth of a new network.* Princeton University Press.

Lehdonvirta, V. (2022). *Cloud empires: How digital platforms are overtaking the state and how we can regain control.* MIT Press. https://doi.org/10.7551/mitpress/14219.001.0001

Lingel, J. (2020). *An internet for the people: The politics and promise of Craigslist.* Princeton University Press. https://doi.org/10.1515/9780691199887

Mauss, M. (2002). *The gift: The form and reason for exchange in archaic societies.* Routledge.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media.* Yale University Press.

Roberts, S. T. (2022). Content moderation. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of big data* (pp. 211–214). Springer International Publishing. https://doi.org/10.1007/978-3-319-32010-6_44

Whitman, M., & Hamilton, J. O. (2010). *The power of many: Values for success in business and in life.* Crown.

Zittrain, J. (1997). Rise and fall of Sysopdom. *Harvard Journal of Law & Technology, 10*(3), 495–514.

CONTENT MODERATION, FROM ADJUDICATION TO CONSENSUS-BUILDING

Emillie de Keulenaar
University of Copenhagen

Moderation has frequently been characterized as an "impossible task" (Kalev Leetaru, 2018), including, as of late, by platforms themselves (Clayton, 2023; Meta, 2025). One of the fundamental challenges of moderation is indeed a normative "demarcation problem"; that is, determining the boundaries between what is and is not acceptable content invariably becomes a motive of political dispute and profound ruptures. For one, the dramatic expansion of voices online — many of whom did not hold the same

prominence under more editorialised media systems (Munger & Phillips, 2020) — complicates a fair execution of content moderation. This task is further complicated by the diversity of standards for acceptable content, both across global user bases and within individual states with deepening political polarization (Gorwa et al., 2020). Historically, too, the "firewalls" erected to prevent the normalization of certain political discourses, particularly in the form of speech legislation, have become extremely vulnerable to profound public disagreement on the meaning, purpose and applications of speech moderation.

In response to these challenges, platforms have adopted strategies designed to make moderation less conspicuous and more dynamic. One approach has been demotion, a technique veering towards a form of normative agnosticism (de Keulenaar, Magalhães, et al., 2023; Gillespie, 2022). That is: content that is deemed acceptable today, be that by users or key stakeholders, is promoted, while content that becomes problematic tomorrow may be demoted. This dynamic process allows platforms to adjust to shifting norms.

Another approach is the use of crowdsourced moderation systems, such as Community Notes (Wojcik et al., 2022). These systems rely on user consensus to classify content as reliable or unreliable, offensive or benign, the goal being to circumvent the perceived partisanship of centralized moderation decisions by fostering collective agreement. This principle is echoed in "bridging systems" (Ovadya and Thorburn, 2023) and "prosocial platform design" initiatives (Schirch, 2023), which redefine moderation as (conflict) *mediation*: that is, to enable productive or "prosocial" relations among users with different backgrounds, other than just *adjudicating* content.Such approaches may build on the conceptual roots of "moderation" as *discourse management*, which emphasize the facilitation and sustainability of dialogue as a form of moderation in its own right. Moderation, in this view, would consist in creating conditions under which conflict arising from radical (political or other) differences does not devolve into various forms of "harmful content", but sustains more or less "constructive" or "productive" dialogue between users.

The conceptual premise of "bridging" moderation systems is that, ultimately, moderation can only be effective if one prevents "harmful" content "upstream". That is, harmful content is considered the product of "bad design", specifically engagement incentives that promote sensationalised and polarized content (Schirch, 2023). Typically, such systems will attempt to compute values like "consensus", "diversity of perspectives", "pluralism" or "prosociality", the aim being either to organise and recommend (or generate) content under a "prosocial" logic (for example, content that is approved across user divides) (Stray et al., 2021; Stray et al., 2023), *or* to have moderation decisions be the product of a process of deliberation and consensus. This implies mathematically remodelling a variety of conflict mediation, dialogue and consensus-building processes taken from various societal conventions, from conflict mediation to the moderation of live political debates, and implementing them as platform design features.

In response, this paper investigates what it means to extend moderation practices into the domain of mediation or consensus-building. What are the conceptual implications of

this shift? Which stakeholders are involved in this type of moderation, and how do they operationalise notions of "prosociality" into moderation mechanisms?

The analysis will address three core dimensions. The first is the conceptual implications of multiplying the purpose and definition of moderation from adjudication alone, to a form of consensus-building for "civic dialogue". The second is the methodological implications of this shift. What kinds of methods and auditing systems can one consider to study forms of moderation that, as of now, may be recorded in one or another platform blog post (X, 2025; X Community Notes, 2025), academic papers (Wojcik et al., 2022), and platform data dumps (X, 2025)? Here, I propose a few ways to combine policy analysis, content moderation practices (de Keulenaar, Alves dos Santos Junior, et al., 2023) with community notes data, using scraping and archival data. Finally, I assess how different stakeholders, particularly state actors, can contribute to and sustain deliberation about such moderation practices by proposing potential frameworks in which other actors can propose bridging models within a wider "middleware" marketplace for platform design.

Its preliminary findings are that bridging systems — in this case Community Notes — have important limitations: the vast majority of notes "still need ratings", and that tends to be the case, perhaps unsurprisingly, when the topic it addresses is intractable (i.e., fails to reach a consensus for a community note to be posted). Still, where there are limitations in content moderation, there are opportunities for speech moderation legislation to complement them. The norms that the DSA puts forth, for example, are also applicable to civic processes, including models for consensus-building that may inform platform design. The DSA could recommend specific features that complement moderation, in a way perhaps analogous to recommending "traffic road signs" beside basic rules of traffic. These may include (but are not exclusive to):

(a) Data and social provenance features (Weyl et al., 2025) (relevant to moderating disinformation and inflammatory content): mechanisms that trace a post's origin, contextualize a user's post by summarizing its underlying arguments and premises based on user-consented history, or other.

(b) Bridging mechanisms: e.g., ranking systems that may prioritise content that crosses user subgroups — or indeed other "bridging" formulas informed by European consensus-building traditions.

(c) Spaces where users can deliberate and come to agreements about what platforms content moderation ought and ought not to moderate and how, offering feedback for DSA applications.

(d) A multi-stakeholder forum to formulate, develop and test deploy additional mechanisms for platform moderation informed by local European norms and civic processes.

References

Clayton, J. (2023, April 12). Elon Musk BBC interview: Twitter boss on layoffs, misinfo and sleeping in the office. BBC. https://www.bbc.com/news/business-65248196

de Keulenaar, E., Alves dos Santos Junior, M., Magalhães, J. C., Rogers, R., & Ganesh, B. (2023, October). After deplatforming: retracing content moderation effects across platforms. The 24th Annual Conference of the Association of Internet Researchers. AoIR2023, Philadelphia, PA, USA. http://spir.aoir.org

de Keulenaar, E., Magalhães, J. C., & Ganesh, B. (2023). Modulating moderation: a history of objectionability in Twitter moderation practices. Journal of Communication, 73(3), 273–287. https://doi.org/10.1093/joc/jqad015

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. Social Media + Society, 8(3), 20563051221117552. https://doi.org/10.1177/20563051221117552

Kalev Leetaru. (2018). Is Social Media Content Moderation An Impossible Task?https://www.forbes.com/sites/kalevleetaru/2018/09/08/is-social-media-content-moderation-an-impossible-task/

Meta. (2025). Misinformation. Transparency Centre. https://transparency.meta.com/en-gb/policies/community-standards/misinformation/

Ovadya, A., & Thorburn, L. (2023). Bridging Systems: Open Problems for Countering Destructive Divisiveness Across Ranking, Recommenders, and Governance. Knight First Amendment Institute, Online. http://knightcolumbia.org/content/bridging-systems

Schirch, L. (2023). The Case for Designing Tech for Social Cohesion: The Limits of Content Moderation and Tech Regulation (SSRN Scholarly Paper No. 4360807). https://doi.org/10.2139/ssrn.4360807

Stray, J., Iyer, R., & Puig Larrauri, H. (2023). The Algorithmic Management of Polarization and Violence on Social Media. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4429558

Stray, J., Vendrov, I., Nixon, J., Adler, S., & Hadfield-Menell, D. (2021). What are you optimizing for? Aligning Recommender Systems with Human Values (No. arXiv:2107.10939). arXiv. https://doi.org/10.48550/arXiv.2107.10939

Weyl, E. G., Thorburn, L., Keulenaar, E. de, Mchangama, J., Siddarth, D., & Tang, A. (2025). Prosocial Media (No. arXiv:2502.10834). arXiv. https://doi.org/10.48550/arXiv.2502.10834

Wojcik, S., Hilgard, S., Judd, N., Mocanu, D., Ragain, S., Hunzaker, M. B. F., Coleman, K., & Baxter, J. (2022). Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation (No. arXiv:2210.15723). arXiv. https://doi.org/10.48550/arXiv.2210.15723

X. (2025). X Community Notes - Working with the Community Notes data. X. https://communitynotes.x.com/guide/en/under-the-hood/download-data

X Community Notes. (2025). Diversity of perspectives. X. https://communitynotes.x.com/guide/en/contributing/diversity-of-perspectives

CITIZENS BAND RADIO, CB SIMULATOR, AND THE PROBLEM OF VOICE

Anna Gibson
Massachusetts Institute of Technology

This paper maps out a new genealogy of digital content moderation by examining two popular communication technologies in the late 1970s and early 1980s: CB (Citizens Band) radio and CompuServe's CB Simulator, the first commercially available digital chat system in the United States. CB radio was a decentralized, analogue communication technology that eventually fell out of favor because of oversaturation: so many people used it that it became unusable. In contrast, CB Simulator quietly introduced content moderation to sustain itself and outlasted the "CB craze." I argue that this case demonstrates an early example of the utility of digital content moderation in contrast with analogous analogue technologies because moderation affords selective voice; drawing from the phrasing of Annany (2018), while communication technology affords the ability to speak, content moderation affords the ability to hear.

It is difficult to emphasize the degree to which CB radio was once culturally important in the United States. In an article in the *Journal of Popular Culture*, Richard Ramsey describes the proliferation of the Citizens Band radio as "the foremost electronic-media phenomenon of the 1970s, rivalling the impact which other technological innovations-the telephone and the television have had on American life" (1979, 340). The technology came to the foreground of American consciousness through the publicity of trucking culture in the late 1970s when U.S. oil prices spiked, and in response the government imposed restrictions on gasoline usage and a nation-wide speed limit. Commercial truck drivers found the restrictions overly cruel and began to protest, using their CB radios to communicate with one another. When the trucker strikes began in the winter of 1973-1974, CB radio was used to coordinate the shutting down of major highways and blocking roads and fuel stations (Hamilton 2008, 217). Truckers also used the CB radio technology to alert other drivers to the presence of police cars with radar and weight stations ("Report …." 1975, 6). These protests and their associated technology became media sensations, appearing on television to demonstrate "how they used the equipment to communicate and organize their forces" during the strikes ("Report …." 1975, 6). Soon truckers became prominent cultural figures themselves as a kind of American cowboy archetype: masculine, white, anti-liberal antiheroes who valued individual freedom above all (Stern 1975). Hollywood released films like *Smokey and the Bandit* (1977) and *Convoy* (1978), in which handsome truckers outwit cops, that performed well at the box office (IMDb).

Key to their tactics was the technology of the lightweight CB radio that afforded decentralized, anonymous, and heavily coded communication. The American population started buying up their own sets to participate, and the growth was exponential. "It took 16 years to get the first million [CB] license applications," according to an FCC engineer, one year for the second million, ten months for the third million, and by February of 1976 there were approximately 200,000 applications a week (Maxwell 1976).

Today, CompuServe is just another ISP owned by AOL, but in the late 80s and early 90s, it was the way many people got online; CompuServe Information Service allowed individuals with personal and terminal computers to access its databases (Poor 2004). It claims it was the first company to offer email to personal computers in 1979, as well as the first to offer "real-time chat online" through its CB Simulator program. Sandy Trevor, former CTO of CompuServe, describes writing the code for CB Simulator over one weekend on the kitchen table (Phelps 1986, 4). One of his technical teams had developed a feature that allowed computer memory to be shared between two users on the same machine, and Trevor saw this as an opportunity to introduce a new service for CompuServe subscribers (personal interview). To design this new chatting service, Trevor chose CB as the model for his new chat program because a lot of people knew about the CB radio; the large scale of the CB radio "craze" meant that there would be widespread understanding of how such a system would operate and be used (personal interview).

Just like CB radio, the simulator used the radio frequency model to create different spaces for users to chat. CB Simulator users, or CBers, could "tune" between 40 channels or use the "scramble" command to have private 1-on-1 discussions. In addition to flipping between channels, users could also tune into several at once with a /MONITOR command. But unlike radio, this does not mean that a user tuning in could do so invisibly. The CB radio had a function to show how many people are currently also in the same channel: the /STATUS command; they describe this as a way to "monitor the heavy traffic areas" (TODAY, 9).

The CB Simulator became so popular it eventually accounted for up 20% of time spent on CompuServe (Tweney 2009). Newspaper articles from the time describe going on CompuServe's CB as a pastime; an LA Times article from 1983 says that the nightly meetings on CB-simulator "are typically referred to as electronic cocktail parties." Users describe the experience as addictive; they go on every night and have every kind of discussion. One user, Cathy Anderson, reported her feelings: "It's very addicting. It gets to be a habit; you live and breathe the things" (McMillan 1981). The article in which she is quoted describes CB simulator as a space for "lonely teenagers" and others who needed someone to talk to. There was even a gay-themed channel, Channel 33, which seems to have existed since at least 1980 as a space for gay community (Poor 2004, 136). There are also records of several marriages that began as discussion on CB Simulator (McMillan 1981; Armstrong 1983).

However, the seemingly endless freedom of CB Simulator was actually quietly matched by back-end moderators who policed the forums. In a New York Times article from 1983, a spokesperson for CompuServe said that "[we] don't censor anybody … but chronic abusers are warned that they will be drummed out of the service unless they

desist" (Lindsey 1983). Sandy Trevor also described the need to prevent pedophiles from harassing teenage users, with several employees having the job of monitoring and preventing such behavior (personal interview). According to a contemporary article in the Guardian, by 1986 there was a language filter overlaid on public chat, and offenders were suspended and required to pay a ten dollar fine (Gold 1986). Such moderation was not prominently advertised in company publications, but effectively it meant that — unlike CB radio — CB Simulator made it through the 1980s. By the end of the 1970s, so many people were using CB radio that there was an overall clogging of the airwaves and interfering with other signals. In the official FCC report, the organization "characterizes CB violations as rampant and CB's overall situation as chaotic" ("Report …" 1975, 3). Users encountered radio channels so full of speech to prevent any kind of meaningful exchange, and there was no effective way to prevent people from buying radios and joining in with vile or awful language. Today we might describe this as an issue of too much speech. CB Simulator, though, was not beholden to such analogue constraints, and the top-down, digital technology allowed moderators to "drum" abusers off the service.

Such moderation, in fact, helped set the stage for modern trust and safety practices. Patricia Phelps, the head sysop for CB Simulator, created documents and conducted public outreach that emphasized logics of care and community in this digital chat space (Phelps 1986). Thus, this cultural moment shows the importance of content moderation in creating spaces for speech that could actually be heard.

References

Ananny, M. (2018). *Networked press freedom: Creating infrastructures for a public right to hear.* MIT Press.

Armstrong, B. (1983, September 9). Innovations: With this modem, I thee wed. *The Washington Post.* ProQuest Historical Newspapers.

Gold, S. (1986, January 23). Computer Guardian: American information dream, available at only 10p a minute / The US CompuServe bulletin database. *The Guardian (London).* LexisNexis: The Guardian.

Hamilton, S. (2008). *Trucking country: The road to America's Wal-Mart economy.* Princeton University Press.

IMDb. (n.d.). Feature film released between 1977-01-01 and 1977-12-31 (Sorted by US box office descending). *IMDb.* Retrieved from http://www.imdb.com

Lindsey, R. (1983, December 2). Computer as letterbox, singles bar and seminar. *The New York Times.* ProQuest Historical Newspapers.

Maxwell, E. (1976, February 16). CB radio: Nation takes to air waves, fer sure, fer sure. *Los Angeles Times.* ProQuest Historical Newspapers.

McMillan, P. (1981, December 6). Hobbyists grasp at new ways to tickle electronic keyboards. *Los Angeles Times.*ProQuest Historical Newspapers: Los Angeles Times, p. B1.

N.A. (1981, July). *TODAY Magazine.* CompuServe Inc.

Phelps, P. (1986). Let's talk!!! An introduction to CompuServe's Citizens' Band Simulator. *CompuServe.*

Poor, N. D. (2004). *Democratic technologies: Openness, decentralization, and the success of information systems*(Doctoral dissertation). University of Michigan.

Ramsey, R. D. (1979). The people versus Smokey Bear: Metaphor, argot, and CB radio. *The Journal of Popular Culture, 13*(2), 338-344.

Report to the Congress by the Comptroller General of the United States. (1975, October 14). *Actions taken or needed to curb widespread abuse of the Citizens Band Radio Service.* Federal Communications Commission, General Accounting Office. GGD-75-88.

Stern, J. (1975). *Trucker: A portrait of the last American cowboy.* McGraw-Hill.

Trevor, S. (2018, October 29). Personal interview.

Tweney, D. (2009, September 24). Sept. 24, 1979: First online service for consumers debuts. *Wired.*https://www.wired.com/2009/09/0924compuserve-launches

THE PAST AND FUTURE OF SPEECH RESTRICTIONS: MEDIA AND PLATFORMIZED SPEECH GOVERNANCE IN INDONESIA AND PAKISTAN

Diyi Liu
University of Copenhagen

Global platform governance operates at the intersection of multiple normative frameworks (Jørgensen, 2017; Haggart & Keller, 2021; Gorwa, 2024). While research has established that social media platforms' market expansion confronts deeply rooted social and cultural speech norms, a clear gap exists in understanding how the governance of online speech has been legitimized in less-than-democratic developing countries (Liu, 2024). Most of these countries lack home-grown tech platforms yet possess large consumer bases, and states have historically maintained strong control over media and speech, even as their democratic institutions evolve. Amid political and economic reforms and increasing internet adoption, civil society organizations have emerged as crucial mediators between global corporate interests and local cultural values.

This paper builds upon a rich tradition of political-economic analysis of media systems. Through document analysis, fieldwork, and expert interviews, it traces the institutional

landscape of state-market-civil society nexus from traditional media to social media governance in Indonesia and Pakistan, examining how different actors have historically negotiated their roles in shaping the boundaries of acceptable (online) speech. The analysis focuses on three interrelated aspects of speech governance: *state regulatory framework*: the evolution from traditional media laws to platform-specific regulations; *market dynamics*: how commercial imperatives shape content governance, considering both domestic media industry interests and global platform companies; and *civil society participation*: how non-governmental organizations, advocacy groups, and other actors shape and contest governance arrangements. This tripartite framework allows us to trace how governance relationships have evolved from traditional media to the platform era, highlighting both continuities and changes in institutional arrangements and power relations.

Indonesia and Pakistan represent compelling cases for examining platformized speech governance for several reasons: Indonesia, the world's fourth-largest social media market, has 139 million users (75% of internet users actively engage with social media; 60% rely on platforms as primary news sources) (Data Reportal, 2024; Newman, 2024). Pakistan is home to 72 million social media users (28% of its population) (Pakistan Telecommunication Authority, 2024). As the world's two most populous Muslim-majority countries, both nations rank among the top for content removal and government takedown requests globally, demonstrating distinct religious influences on their political and cultural fabric that shape content moderation approaches (Lombardi, 2013; Younas et al., 2020). In Indonesia, high social media dependence has raised concerns about online harassment and hate speech (Lim, 2017), particularly during the 2024 presidential election when platforms amplified tensions for marginalized groups (Heychael et al., 2024). In Pakistan, platforms like TikTok face persistent criticism as "indecent and immoral" from state, courts, and the public, occasionally resulting in platform bans and even violence stemming from content controversies (Ng, 2025).

Historical Context of Speech Governance: Tracing the Evolution

In Indonesia, the transition from Suharto's authoritarian New Order regime (1966-1998) to post-*Reformasi* democratization offers crucial insights into evolving speech governance. After 1998, Indonesia experienced rapid press freedom expansion and media marketization (Mulya Lubis, 2017; Nugroho et al., 2012), followed by concentration into politically-connected media conglomerates (Tapsell, 2015). Two key legislative developments shaped this landscape: the Press Law (No. 40/1999), which established an independent Press Council with authority to self-regulate the industry (Article 15(2)(f)), and the Broadcasting Law (No. 32/2002), which created the Indonesian Broadcasting Commission and prohibited content related to slander, violence, obscenity, and content polarizing ethnicity, religion, race, and inter-group relations (SARA). Similar speech restrictions are encoded into regulations for digital speech platforms. The Electronic Information and Transaction Law (*UU ITE*) and its implementing regulations (No. 71/2019, No. 5/2020) imposed strict content removal timeframes while granting broad ministerial discretion, raising due process concerns (Audrine & Setiawan, 2021). The 2024 amendment of the ITE Law further narrowed legal definitions while expanding jurisdictional reach over international platforms, revealing tensions between democratic aspirations and traditional control impulses.

Pakistan followed a rather different trajectory, with state monopoly dominating until the early 1990s (Gul et al., 2017). The paradoxical turning point came during General Musharraf's military regime (1999-2008) with the 2002 media liberalization (Sulehria, 2017), which established the Pakistan Electronic Media Regulatory Authority (PEMRA) through a two-tiered system: government-appointed Councils of Complaints and the government-controlled PEMRA itself (Digital Rights Foundation, 2023). Unlike the self-regulatory model of Indonesia's Press Council, Pakistan's Press Council operated under explicit subordination to "national interest", with regulations prohibiting content related to sectarianism or material deemed prejudicial to Pakistan's ideology (Imran, 2022). Social media governance falls under the Pakistan Telecommunication Authority (PTA), with continuing military influence through leadership appointments. The Prevention of Electronic Crimes Act (2016) and Social Media Rules (2021) require platforms to remove broadly defined "unlawful" content within 48 hours, including material offending the "glory of Islam," public order, or morality. Recent expansions of state control through the E-Safety Bill (2023) and 2025 amendments to PECA (Guramani, 2025) further criminalize "fake news" while establishing new regulatory authorities specifically for social media platforms, demonstrating how seemingly democratic reforms can reinforce state control over expression in the platform era.

Implications for Platform Governance

The comparative historical analysis seeks to reveal how media governance approaches continue to shape contemporary platform regulation in morally conservative contexts. Both countries demonstrate persistent tensions between democratic aspirations, state control impulses, moral and religious fabric, as well as commercial interests. The examination of these two countries' experiences contributes to broader debates about freedom of expression in the digital age, particularly in contexts where democratic institutions are still evolving.
In both cases, we observe how regulatory logics from traditional media governance extend into the platform era, with historical patterns of state control taking new forms in digital contexts. Indonesia's self-regulatory press model offers potential institutional arrangements for platform governance, though its implementation has been inconsistent. Pakistan's military-influenced liberalization shows how state strengthens control over expression, increasingly on the digital infrastructure level. Understanding these historical trajectories and contextual factors is essential for developing more nuanced approaches to platform governance that acknowledge the complex interplay between local speech norms, state regulatory frameworks, and global platform policies. Such insights can inform both scholarly debates and practical governance models for societies navigating the challenges of digital transformation while balancing religious sensitivities, political interests, and democratic expression.

References

Audrine, P., & Setiawan, I. (2021). *Impact of Indonesia's Content Moderation Regulation on Freedom of Expression*. Center for Indonesian Policy Studies.
https://doi.org/10.35497/347642

DataReportal. (2024, February 21). *Digital 2024: Indonesia*. DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2024-indonesia

Digital Rights Foundation. (2023). *Policy Paper: Analyzing Content Regulation and Model of the Council of Complaints of the Pakistan Electronic Media Regulatory Authority (PEMRA) in light of the Civil Petition No.3506 of 2020, Pakistan Electronic Media Regulatory Authority (PEMRA) & another Versus M/s ARY Communications Private Limited (ARY Digital)*. https://digitalrightsfoundation.pk/wp-content/uploads/2023/04/FINALVERSION_14-04-2023_Content-regulation-in-Pakistan-and-the-model-of-the-Council-of-Complaints-of-Pakistan-Electronic-Media.pdf#page=25.58

Gorwa, R. (2024). *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*(1st ed.). Oxford University PressNew York. https://doi.org/10.1093/oso/9780197692851.001.0001

Gul, M., Obaid, Z., & Ali, S. (2017). Liberalization of Media in Pakistan: A Challenge to Democracy. *The Journal of Humanities and Social Sciences*, *25*(1), 37–54.

Guramani, N. (2025, January 22). *Latest Peca changes draft lowers jail term to 3 years, fines remain at Rs2m*. DAWN.COM. https://www.dawn.com/news/1886993

Haggart, B., & Keller, C. I. (2021). Democratic legitimacy in global platform governance. *Telecommunications Policy*, *45*(6), 102152. https://doi.org/10.1016/j.telpol.2021.102152

Heychael, M., Arum, N. S., & Nabiyyin, H. (2024). *Freedom or Hate? Assessing Social Media Platform's Accountability for Hate Speech Targeting Vulnerable Groups During 2024 Indonesia Election*. Southeast Asia Freedom of Expression Network (SAFEnet).

Imran, S. (2022). PEMRA's Overreach: Exploring Censorship and State Anxiety. *Reel Pakistan: A Screen Studies Forum*, *4*, 90–99.

Jørgensen, R. F. (2017). What Platforms Mean When They Talk About Human Rights. *Policy & Internet*, *9*(3), 280–296. https://doi.org/10.1002/poi3.152

Lim, M. (2017). Freedom to hate: Social media, algorithmic enclaves, and the rise of tribal nationalism in Indonesia. *Critical Asian Studies*, *49*(3), 411–427. https://doi.org/10.1080/14672715.2017.1341188

Liu, D. (2024). Borderline content and platformised speech governance: Mapping TikTok's moderation controversies in South and Southeast Asia. *Policy & Internet*, *16*(3), 543–566. https://doi.org/10.1002/poi3.388

Lombardi, C. B. (2013). Designing Islamic constitutions: Past trends and options for a democratic future. *International Journal of Constitutional Law*, *11*(3), 615–645. https://doi.org/10.1093/icon/mot038

Mulya Lubis, T. (2017). Two Faces of Freedom of the Press in Indonesia's Reformation Era. In D. Weisenhaus & S. N. M. Young (Eds.), *Media law and policy in the internet age* (pp. 227–236). International conference on Media Law and Policy in the Internet Age, Oxford Portland, Oregon. Hart Publishing. https://doi.org/10.5040/9781782257417

Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). *Reuters Institute digital news report 2024*. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-VY6N-4V57

Ng, K. (2025, January 30). Pakistan: US teen shot dead by father over TikTok videos. *BBC News*. https://www.bbc.com/news/articles/cy8pvw3xxxeo

Nugroho, Y., Putri, D. A., & Laksmi, S. (2012). *Mapping the landscape of the media industry in contemporary Indonesia.* (Engaging Media, Empowering Society: Assessing Media Policy and Governance in Indonesia through the Lens of Citizens' Rights). Centre for Innovation Policy and Governance and HIVOS Regional Office Southeast Asia.

Pakistan Telecommunication Authority. (2024). *Annual Report 2024*. https://www.pta.gov.pk/assets/media/2024-12-16-pta_annual_report.pdf

Sulehria, F. (2017). DD and PTV as victims of media globalisation. *Asian Journal of Communication*, *27*(1), 97–112. https://doi.org/10.1080/01292986.2016.1255237

Tapsell, R. (2015). Platform convergence in Indonesia: Challenges and opportunities for media freedom. *Convergence: The International Journal of Research into New Media Technologies*, *21*(2), 182–197. https://doi.org/10.1177/1354856514531527

Younas, F., Naseem, M., & Mustafa, M. (2020). Patriarchy and Social Media: Women Only Facebook Groups as Safe Spaces for Support Seeking in Pakistan. *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, 1–11. https://doi.org/10.1145/3392561.3394639