Yale Journal of Law & Technology
Volume 26, Issue 2

The Platform Federation

Gilad Abiri[*] **& Sebastián Guidi**[**]

We are witnessing the birth of a Platform Federation. Global platforms wield growing power over our public sphere–and yet our politics and public debates remain stubbornly state-based. In the platform age, speech can transcend international boundaries, but the repercussions of speech are mainly felt within our own domiciles, municipalities, and national territories. This mismatch puts countries in a difficult place, in which they must negotiate the tension between steering the public sphere to protect local speech norms and values and the immense benefits of free transboundary communication. This Article explores the outcome of this balancing act—what we call platform federalism: where it comes from, how it is unfolding, and how to make it better.

The rise of global digital platforms brought up a crisis that has not yet been fully diagnosed. Until their appearance, the public sphere was disciplined by gatekeepers such as traditional mass media and other civil society institutions. They acted to enforce a common set of norms over public discourse. These gatekeepers fulfilled crucial social functions. They enacted and enforced the fundamental social norms that made public communication possible, while at the same time avoiding direct state intervention in public discourse. Through social media, people are now able to bypass these institutions and reach mass audiences directly—what we call the "bypass effect."

Countries are reacting to the consequences of the bypass effect by enforcing local social norms directly. Autocracies might enjoy the dubious luxury of shutting down Internet borders completely. This option, however, is not available for democracies, nor is it desirable. Democracies have embraced

---

[*] Assistant Professor of Law, Peking University School of Transnational Law and Affiliate Fellow, Information Society Project, Yale Law School.
[**] Lecturer, Universidad Torcuato Di Tella and Visiting Fellow, Information Society Project, Yale Law School.

softer forms of regulation, which we call "state federalism." As civil-society gatekeepers are bypassed, states take the mission of curating the public sphere onto themselves: they forcefully impose their own civility norms on platforms' users (like Germany) or directly forbid fake news on them (like France). State federalism might work in restoring the public sphere's civility, but it risks unduly imposing the state's (as opposed to the community's) values upon the population. State federalism, in other words, can quickly become incompatible with liberalism.

We propose a new set of policy tools to maintain domestic civility in the public sphere while keeping state power at bay: civil society federalism. In civil society federalism, the state does not police the public sphere by itself, but rather requires platforms to invite civil society back into their gatekeeping role. These policies ask civil-society organizations to shape the norms that constitute public discourse; as in the past, they are the ones to exclude hate speech, profanity, or misinformation from the public sphere. By bringing civil society back, states can ensure the civility of the public sphere without exerting undue power over it.

Article Contents

Introduction

The early Internet kindled the vision of a world without borders.[1] Soon after, emerging social-media platforms nurtured the promise of a "global community,"[2] a "global public consciousness."[3] With enough time, it was hoped, this new technology could bring about the vision of a global democracy.[4] Indeed, on social media we are able to debate abortion with a religious woman in Sri Lanka, or affirmative action with a Turkish immigrant in Germany.

However, borderless communication has led to power centralization. Influence, once dispersed among local civil societies within nations, has consolidated within a handful of digital platforms. This shift has given rise to a novel breed of digital gatekeepers that lie beyond political borders.[5]

Thus arose one of the most profound conflicts embedded within the fabric of the Internet: although our public sphere is now dominated by global digital platforms, our politics remains stubbornly local.[6] Most issues, even those presenting scale

---

[1] The most famous example of the early Internet's cosmopolitan dreams is John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), https://www.eff.org/cyberspace-independence [https://perma.cc/QHR9-BMKQ].

[2] Mark Zuckerberg, *Building Global Community*, FACEBOOK (Feb. 16, 2017), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634 [https://perma.cc/AMJ4-9MZH].

[3] This is the definition of Twitter given by its founder, Jack Dorsey: Twitter is "the closest thing we have to a global consciousness," Jack Dorsey (@jack), TWITTER (Apr. 26, 2022, 10:03 PM), https://twitter.com/jack/status/1518772753460998145 [https://perma.cc/F6MT-TBQ2].

[4] *See e.g.*, Hans Klein, *The Feasibility of Global Democracy: Understanding ICANN's At-Large Election*, 3 INFO. 333, 334 (2001) ("[T]he problems inherent in global governance are partially counteracted by new communication technology, most notably the internet. The internet makes global person-to-person communication possible, and this, too, helps surmount barriers to political community.").

[5] One of us has defined this phenomenon as one of "disconnected gatekeepers." *See* Gilad Abiri, *Moderating from Nowhere*, 47 BYU L. REV. 757, 797-99 (2022).

[6] Angelos Chryssogelos, *Populism Is Not a Bug but a Feature of Globalization*, NOEMA MAG. (July 30, 2020), https://www.noemamag.com/populism-is-not-a-bug-but-a-feature-of-

dilemmas or externalities, are typically deliberated and determined at a national level.[7] While interactions with overseas debates can provide new perspectives to integrate into local public discourse, the rise of foreign election interference as a primary public concern[8] underscores an increasingly vital challenge: how can these global entities fit our locally established forms of trust? And, more broadly, how can we ensure that our public debates take place on trustworthy and trusted digital platforms?

The governments of the world reacted to the dominance of global digital platforms in different ways, sometimes clever, sometimes clumsy. Governments—democracies and autocracies alike—block foreign platforms, impose stringent local laws, and bolster their indigenous infrastructure and hardware.[9] Yet, beneath this fragmentation, most countries are not seeking a complete split—a "Splinternet," as some have

---

globalization [https://perma.cc/QSQ9-7GQK] ("While economic exchanges and social, cultural and environmental trends and flows have indeed become global in scope in the last 30 years, politics has remained stubbornly anchored to settings defined by national borders and sovereignty.").

[7] For the conceptual problems posed by domestic decisions that cause externalities, *see generally* Sebastián Guidi & Nahuel Maisley, *Who Should Pay for COVID-19? The Inescapable Normativity of International Law*, 96 N.Y.U. L. REV. 375, 419-27 (2021) (explaining the inherent difficulty in assessing how countries harm each other, since one must rely on accounts of "harm" and "causation" that are deeply affected by local values that differ among different communities).

[8] Jens David Ohlin & Duncan B. Hollis, *Introduction*, *in* DEFENDING DEMOCRACIES: COMBATING FOREIGN ELECTION INTERFERENCE IN A DIGITAL AGE 1, 1 (Jens David Ohlin & Duncan B. Hollis eds., 2021) ("Election interference is one of the most widely discussed international phenomena of the last five years.").

[9] *See generally* Anupam Chander & Haochen Sun, *Sovereignty 2.0*, 55 VAND. J. TRANSNAT'L L. 283, 292 (2022) (describing the achievement of digital sovereignty as a central aspiration of states); Mark A. Lemley, *The Splinternet*, 70 DUKE L.J. 1397 (2021) (describing this general tendency in which "what you see is actually an ongoing nation-by-nation competition for who gets the internet"). *See also id.* at 1408 ("[I]t's not just authoritarian governments that are using balkanization to lock down the internet.").

called it.[10] Instead, they are actively seeking to regain some control over their own public sphere while remaining connected to the global-platform sphere. Most actors are not driven to erect insurmountable "walled gardens;"[11] rather, they aspire to cultivate and shape their own "hedges." We are not witnessing the birth of the Splinternet, we posit, but of a Global Platform Federation.

This Article defines and examines the emerging phenomenon of platform federalism: where it comes from, how it is unfolding, and how to make it better. By adopting platform federalism, we argue, both countries and platforms benefit. States, on one hand, will be able to preserve the benefits of cross-border communication without unduly encroaching on freedom of speech. For platforms, on the other hand, platform federalism represents a way to address the deep legitimacy crisis arising from platforms' attempts to exercise state-like decision power through processes that formally resemble law but have no connection to the people from which actual law draws authority.[12]

Platform federalism resists simplistic reductions to either of two archetypal perspectives on internet governance. On one end, there are those who still harbor hopes of a global Internet, unbothered by national or regional borders. Whether they are moved by cosmopolitan drives or by imperialistic ambitions, these Internet "universalists" see globality in Internet communication as an "unqualified human good."[13] On the other end, Internet "localists" are afraid of foreign influences

---

[10] *Id.* at 1418-19 ("[W]e're losing the internet. We're replacing it with 'the splinternet,' a balkanized set of computer protocols that increasingly differs by company and by country.").

[11] *Id.* at 1399 ("[T]he internet is being balkanized. We are returning to walled gardens. Some of those walled gardens are run by private companies, but increasingly, they are being created by drawing national boundaries around the internet.").

[12] *See* Gilad Abiri & Sebastián Guidi, *From a Network to a Dilemma: The Legitimacy of Social Media*, 26 STAN. TECH. L. REV. 92, 135-41 (2023). By making content moderation responsive to the particular norms of each community, platform federalism has the potential of enhancing the legitimacy of platforms' decisions.

[13] We are paraphrasing E.P. Thompson's famous contention about the rule of law. *See* Morton J. Horwitz, *Review: The Rule of Law: An Unqualified Human Good?*, 86 YALE L.J. 561, 566 (1977).

in their social life and public culture. Therefore, they attempt to shut down the borders of their own, local platform Internet.

The *first* goal of this Article is to describe this spectrum and show that most countries do not fall into either of these extremes. Most countries want to stay on the global Internet, but leave at the same time,[14] creating a model of internet governance similar to what comparative constitutional law would call "federalism." They are eager to reap the cultural and economic benefits of globalization, but they also fear the corrosive effects of global social media on their public sphere. They are not necessarily atavistic or autocratic, but they give a great deal of importance to the fact that, "[i]f the institutions of democracy and capitalism are to work properly, they must coexist with certain premodern cultural habits that ensure their proper functioning."[15] The more multinational companies—with no ties and no accountability to their local circumstances—govern the domestic public sphere, the less governments feel they can maintain the invisible ties that keep their communities together. Smaller communities (say, nation states) are perpetually negotiating the terms of engagement with the broader communities to which they belong (say, metaphorically, digital platforms operating across national borders).[16] In federal arrangements, smaller communities retain some powers for themselves in issues they deem central to their communal identities, while delegating other decisions to a central, distant source of power.[17] We suggest countries are adopting a similar type of arrangement when it comes to content moderation on the platform Internet: they tolerate, or even welcome, global platform power over an increasingly crucial digital public sphere, but they retain some powers for themselves so their public sphere does not become

---

[14] *See infra* note 223 and accompanying text.
[15] FRANCIS FUKUYAMA, TRUST: THE SOCIAL VIRTUES AND THE CREATION OF PROSPERITY 11 (1995).
[16] *See infra* notes 221-226 and accompanying text.
[17] *Id.*

unrecognizable.

The *second* goal of this Article is to show that, in adopting different forms of platform federalism, countries are fighting a crisis caused by the rise of global digital platforms: people's increasing ability to bypass the institutions that up to now had acted as gatekeepers of public speech. Understanding the reason states and local actors see this "bypass effect" as a problem in need of a solution requires a deep examination of how pre-digital speech structured public discourse.

Public discourse needs to follow a set of common social norms that distinguish acceptable speech from verboten utterances. Without these norms—that prevent, for instance, hate speech, insults, and misinformation—the public sphere is not the organized deliberation of the community, but a chaotic cacophony.[18] In the pre-digital age, established gatekeepers such as mass media, universities, and other civil-society institutions fulfilled this function, curating and controlling public discourse.[19] These gatekeepers operated to maintain a relatively common set of speech norms in public discourse. Mass-media elites had a tight grip on the ability to broadcast messages to a mass audience: public debate *was* what happened in newspapers and on TV. This feature of mass communication gave gatekeepers the power of enforcing the social norms that maintained social cohesion. For all the problems this concentration of power had, these gatekeepers provided a safeguard against uncontrolled hate speech and senseless fake news.[20]

The Internet came to challenge this arrangement. From the comfort of their homes, citizens worldwide can bypass the institutions that used to act as gatekeepers of public discourse. Foreigners can easily broadcast or diffuse a message to the masses anywhere. Left to its own devices, nothing ensures that discourse will be civil or organized.

Social-media platforms have taken the role of managing public discourse onto themselves, but at great cost for state-

---

[18] For the relationship between civility norms and democratic deliberation, see generally Robert C. Post, *Racist Speech, Democracy, and the First Amendment*, 32 WM. & MARY L. REV. 267, 286-87 (1991). *See also infra* Section II.B.

[19] *See* Abiri, *supra* note 5 at 794-97.

[20] *See id.*

based politics: previously local discourse is now adjudicated by American or Chinese companies operating across varied local contexts worldwide. The gatekeeping role once played by local elites has been outsourced to distant corporate entities.

As a consequence of this change in forum, the nature of control over public discourse has fundamentally shifted in the digital age from gatekeeping to content moderation. In pre-digital times, speech norms were enforced by controlling what information could be published in the first place. A small elite decided what was acceptable for their community, and maintained speech norms by preventing unacceptable content from ever seeing the light of day. However, on digital platforms with billions of users generating massive volumes of content, this approach of pre-publication filtering is infeasible. Instead, platforms are forced to rely on post-publication content moderation, reviewing content against established rules and community standards after the content is posted. The scale of this endeavor, with platforms moderating trillions of posts, is unprecedented.

In this context, it does not come as a surprise that governments have reacted to the destabilization of the pre-digital equilibrium. Predictably, illiberal countries, such as China, have erected powerful firewalls around their digital borders.[21] However, democracies like France, Germany, and Brazil have also enacted their own regulations to secure some form of state control over what can and cannot be said online.[22] Governments have coercively required platforms to take down content they deem illegal, such as uncivil utterances or misleading information. We call this type of reaction, in which the state itself is the one to police the border between the publishable and the unpublishable, "state federalism."

However, this state federalism comes at a cost. In the pre-

---

[21] *See, e.g.*, Daniel Anderson, *Splinternet Behind the Great Firewall of China: Once China Opened Its Door to the World, It Could Not Close It Again*, 10 QUEUE 40 (2012).

[22] *See infra* Section VI.A.

digital age, delegating the enforcement of speech norms to non-state institutions had enabled liberal states to precariously solve an inherent paradox of democratic governance. As any state, the liberal state[23] relies on social cohesion, which in turn depends on the respect for common social norms. However, were the state to enforce these norms directly, it would risk compromising its liberal nature, potentially infringing upon free-speech principles.[24] This is the reason why liberal regimes largely abstained from directly regulating speech and delegated the discipline of the public sphere to civil society institutions, such as mass media and universities, which are independent from governments. With this newfound eagerness to regulate speech online, we are now left to wonder whether we have given up on a fundamental liberal safeguard.

The *third* goal of this Article is therefore normative. It is our contention that nations have good reasons to attempt to mitigate the bypass effect, yet they should attempt to do so without undermining the self-imposed limitations of liberal governance. Governments should adhere to free-speech principles and not directly regulate content-moderation policies. In order to restore the pre-digital liberal equilibrium between freedom of speech and enforcement of social speech norms, states should try what we call "civil society federalism."

Under the civil society federalism model, states and platforms work together to bring civil-society institutions back into the business of managing the public sphere. When it comes to deciding what counts as online hate speech, abuse, or misinformation, it is crucial for decisionmakers to be sensitive to local context. Our proposed strategy involves deliberate regulation requiring platforms to give domestic civil society a meaningful role in curating and structuring public discourse.

Civil society federalism aims at maintaining domestic community norms while avoiding the risk of illiberalism that stems from direct state involvement in speech regulation.

---

[23] Following the influential Rawlsian account, by "liberal state" we refer to one in which power "may only be used in ways that all citizens can reasonably be expected to endorse" independently of their "comprehensive doctrines" about good and bad. *See* Leif Wenar, *John Rawls*, STAN. ENC. PHIL. (Apr. 12, 2021), https://plato.stanford.edu/entries/rawls [https://perma.cc/P7N2-U8UT].

[24] *See infra* Part III.

Furthermore, as this process shifts from organic to deliberate, it presents an opportunity for communities to actively participate in selecting which civil-society organizations best represent their diverse societal voices in an equitable way.

We propose that states should adopt legislation mandating the inclusion of local civil-society organizations in three key areas, with the aim of restoring their role as gatekeepers of information: enforcing content moderation, implementing policy, and establishing common trusted sources of information.

This Article begins with the premise that platforms are managed at a global level, while most of our politics are carried out at a domestic level. Part I suggests that dealing with this fundamental tension is necessary if social media platforms are going to become "trusted and trustworthy" institutions in our public cultures. Part II explores the relationship between speech and community, and argues that our politics need some form of domestic control over its own social norms. Part III shows that this need has traditionally put liberal states in tension: they want to ensure solidarity ties within their communities, but cannot do so through force without relinquishing their liberal character. Part IV shows how pre-digital societies dealt with this problem by establishing non-state institutions in charge of gatekeeping the public sphere, and how the advent of the platform Internet challenges this arrangement. It then analyzes state responses to the "bypass effect" and stylizes them around two poles: universalism and localism. Part V explores platform universalism and localism in greater detail. Finally, Part VI introduces state federalism and civil society federalism.

I. A (Local) Trusted and Trustworthy Public Sphere

Social media platforms dominate public spheres across the globe. Once celebrated, the advent of platform-based speech is now seen as responsible for many of our current social woes,

including the rapid spread of hate speech and misinformation.[25] Some scholars, however, see these issues as symptoms of a more fundamental disorder—the fact that in the age of digital platforms, as Jack Balkin puts it, we lack "trusted and trustworthy organizations for facilitating, organizing, and curating public discourse."[26] Without such institutions and professions, any public sphere "will decay . . . [, w]eaken the institutions or destroy trust, and the public sphere becomes a rhetorical war of all against all, where no one is believed except the members of one's own tribe, and people cleave to whatever beliefs are most comforting to them."[27] Without trust in the institutions that are meant to tell us what is reliable knowledge or which utterances fall beyond the pale of public discourse, we are left in a free-for-all that undermines fundamental free-speech values, be they political self-government, cultural democracy, or the ability of society to produce common knowledge.

The question, then, is how to produce trusted and trustworthy intermediate institutions in the digital platform ecosystem. Designing a remedy depends on how we understand the causes of the current environment of mistrust.

Balkin sees the problem of diminishing trust as arising from "informational capitalism."[28] The reason we do not trust social-media platforms is because they engage in what Shoshana Zuboff calls "surveillance capitalism"—the ad-based monetization of personal information through the collection and processing of personal data.[29] Surveillance capitalism undermines trust in various ways. First, because it leads platforms to seek the maximization of engagement,

---

[25] In a previous article, we exemplified this fall from grace with the very different messages about social media brought by the films *The Social Network* and *The Social Dilemma*. *See* Abiri & Guidi *supra* note 12, at 94.

[26] Jack M. Balkin, *To Reform Social Media, Reform Informational Capitalism*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 233, 234 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022).

[27] *Id.* at 242.

[28] *Id.*

[29] *See generally* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019) (coining and defining the phenomenon of surveillance capitalism).

surveillance capitalism creates incentives to promote material that produces strong emotions, "even if some of that material turns out to be false, misleading, undermines trust in knowledge-producing institutions, incites violence, or destabilizes democracies."[30] Second, as the model requires massive data collection, platforms have little incentive to protect users' privacy and to educate them about what is done with the data collected about them. Finally, "[b]ecause social media companies do not fully internalize the social costs of their activities, they will tend to skimp on content moderation that does not increase their profits."[31] To correct these problems, Balkin turns to different market-based proposals, including antitrust law, information fiduciaries, and tweaking of the current liability-shield regime.[32]

However, this characterization of the problem leaves a defining feature of platforms unexplored. As a perhaps incidental consequence of their business model, global platforms remain crucial actors in how the public sphere is disciplined. They curate content, indicate fake news, and censor hate speech and otherwise uncivil utterances. Powerful social media platforms, therefore, shape the public sphere in ways similar to those of mass media.[33]

When combined with the global reach of platforms, this novelty in mass communication brings us to a fundamental problem. While platforms are global, and so is their governance, the most relevant site for deliberation and formation of public opinion remains the national public sphere. Most political decisions are still made at the level of nation-states,[34] most people still define their identities primarily in

---

[30] Balkin, *supra* note 26, at 243.
[31] *Id.* at 244.
[32] *Id.* at 245-53.
[33] *See* discussion in *infra,* Section IV.B.
[34] Even the European Union, the most advanced regional integration project worldwide, lacks power in a wide variety of crucial issues, such as tax policy, social security, and defense. It is commonplace to assert that this

national terms,[35] and most media is produced and consumed within national borders.[36] Despite what they may tell us, global social-media platforms are not custodians of a *global* public sphere that has not yet fully emerged.[37] They are, rather, intervening in hundreds of *local* public spheres, at the same time, and with the same rules.

The global character of social media, therefore, poses additional challenges to those identified by Balkin and others. It is true we need social-media platforms to start operating in a way that is seen as promoting the public good so they can be

---

lack of a European demos is fatal for European democracy. *See e.g.*, Fritz W. Scharpf, *Legitimacy in the Multilevel European Polity*, 1 EUR. POL. SCI. REV. 173, 177 (2009) ("[As] the peoples of 27 member states lack a collective identity that could legitimate Europe-wide majority rule . . . they would presently lack all the societal and institutional prerequisites of input-oriented democracy: No Europe-wide media of communication and political debates, no Europe-wide political parties, no Europe-wide party competition focused on highly salient European policy choices, and no politically accountable European government that must anticipate and respond to the egalitarian control of Europe-wide election returns.").

[35] *See* Asier H. Aguirresarobe, *Is National Identity in Crisis? An Assessment of National Imaginations in the Early 2020s*, 22 STUD. ETHNICITY & NATIONALISM 14, 15, 24 (2022) (arguing that despite globalizing trends "nationalism remains almost hegemonic as a form of political and communal identification," and therefore it is "difficult to imagine any trend towards globalization that is not driven by national ideas").

[36] Terry Flew, Petros Iosifidis & Jeanette Steemers, *Global Media and National Policies: The Return of the State*, *in* GLOBAL MEDIA AND NATIONAL POLICIES: THE RETURN OF THE STATE 9 (Terry Flew, Petros Iosifidis & Jeanette Steemers eds., 2016) ("[M]ainly national media systems are still rooted in geographically located national media cultures and domestic politics and power. Preferences for local content remain strong . . . , particularly in television/screen media where the nation-state is still the main location through which the media address their audiences. Media industries still mostly think in terms of national and regional markets . . . .").

[37] Nancy Fraser, *Transnational Public Sphere: Transnationalizing the Public Sphere: On the Legitimacy and Efficacy of Public Opinion in a Post-Westphalian World*, 24 THEORY, CULTURE & SOC'Y 7, 10 (2007) ("At least since its 1962 adumbration by Jürgen Habermas, public sphere theory has been implicitly informed by a Westphalian political imaginary: it has tacitly assumed the frame of a bounded political community with its own territorial state. . . . Is the concept of the public sphere so thoroughly Westphalian in its deep conceptual structure as to be unsalvageable as a critical tool for theorizing the present? Or can the concept be reconstructed to suit a post-Westphalian frame?").

trusted. However, existing forms of political trust and legitimacy are deeply embedded in our current form of politics: democratic nation-states.[38] What distinguishes Facebook, Reddit, Threads, or X from Le Monde, Asahi Shinbun, or the BBC is not only their business model or their technological capabilities but also that they are U.S. corporations managed by U.S. elites. The fact that platforms operate across borders means that they will inherently struggle to fit into locally established forms of trust. Being detached from any particular public sphere, they are also unable to reconstitute them.

To fully grasp this problem, we now examine the way public spheres are constituted and regulated. Only then we will be able to imagine the ways in which social-media platforms can contribute to their strengthening.

II. Speech Norms and Democracy

Social media fundamentally disrupts the ways we communicate to each other, both in private and in public.[39] In this Article, we argue that one of the fundamental ways in which this disruption happens is through disempowering the gatekeepers of social discourse, in a process we will call the "bypass effect."[40]

As our societies are built fundamentally through discourse,[41] fundamentally altering the ways we communicate to each other is bound to have momentous consequences for

---

[38] *See id.* at 13 ("In this model, democracy requires the generation, through territorially bounded processes of public communication, conducted in the national language and relayed through the national media, of a body of national public opinion. This opinion should reflect the general interest of the national citizenry concerning the organization of their territorially bounded common life, especially the national economy. The model also requires the mobilization of public opinion as a political force.").

[39] *See generally* LEE RAINIE & BARRY WELLMAN, NETWORKED: THE NEW SOCIAL OPERATING SYSTEM (2012) (analyzing the ways in which social media has changed our communication practices).

[40] *See infra* Section II.A.

[41] *See infra* Section II.B.

the very constitution of the political community: a community is what it is, in part, because of the way it talks.[42] When we change the way we talk to each other, we also change who we are.[43]

To understand the challenges that social-media platforms pose to democracy, we need to explore the links between political community, speech, and democracy. This Part explores these linkages.

A. Community

A democracy requires a *demos*, a bounded community that is both the subject and the object of self-government.[44] Many authors discard the very possibility of global democracy precisely because of the impossibility of constituting a global demos.[45] Prominent authors are even skeptical of the possibility of building a European democracy within the institutional framework of the European Union, because of the lack of a European demos that can provide the necessary sociological ground for such a project.[46] Democracies have hitherto been reliant on "substantive values and principles,"[47] and therefore, every "modern democratic state demands a 'people' with a strong collective identity."[48]

The idea that a democracy must rely on some form of community is implicit in the very idea of self-rule. As Charles Taylor put it, if "the people [are] supposed to rule; this means

---

[42] *See infra* Section II.C.

[43] *See, e.g.*, SHERRY TURKLE, RECLAIMING CONVERSATION: THE POWER OF TALK IN A DIGITAL AGE (2015) (generally arguing that digital age communication modifications have consequential impacts on our personal and collective identities).

[44] *See, e.g.*, JOHN DUNN, BREAKING DEMOCRACY'S SPELL 24 (2014) ("Democracy requires a demos. It requires one conceptually if it is to be a coherent idea. It requires one semantically if it is to be a noncontradictory description of a possible historical segment of the human world. More formidably, it requires one entirely practically if it is to come into existence and persist over time in the world.").

[45] *See e.g.*, Sarah Song, *The Boundary Problem in Democratic Theory: Why the Demos Should Be Bounded by the State*, 4 INT'L THEORY 39 (2012) (rejecting the idea of a global demos and a global state as infeasible).

[46] *See supra* note 34 and accompanying text.

[47] Song *supra* note 45, at 41.

[48] Charles Taylor, *Democratic Exclusion (and Its Remedies?)*, *in* DILEMMAS AND CONNECTIONS: SELECTED ESSAYS 124, 130 (2011).

that the members of this 'people' make up a decision-making unit, a body which takes joint decisions."[49] The concept of self-rule is elusive in the absence of a political community. If a sense of collective identity is not ingrained in the political culture, the question arises as to who exactly would be the self-governing entity. Decisions are not made by an impersonal algorithm that aggregates preferences; rather, they are made by "We the People."[50]

The political identity of a self-governing community is not only a conceptual byproduct but a functional need. Engagement in public discourse and self-rule requires the active participation and dedication of citizens. It demands that we take time to form an opinion on public issues, that we invest our energy in discussing it with other people, that we risk engaging in fights with acquaintances, and that we dedicate hours to standing in line in polling stations.[51] This level of commitment is made possible by the existence of a shared political community. In fact, a "citizen democracy can only work if most of its members are convinced that their political society is a common venture of considerable moment and believe it to be of such vital importance that they participate in the ways they must to keep it functioning as a democracy."[52] To put it succinctly, democratic liberal states, which impose considerable obligations upon their citizens, require "much greater solidarity toward compatriots than toward humanity in

---

[49] *Id.* at 129.

[50] For a critique of aggregative theories of collective will formation, see Richard H. Pildes & Elizabeth S. Anderson, *Slinging Arrows at Democracy: Social Choice Theory, Value Pluralism, and Democratic Politics*, 90 COLUM. L. REV. 2121, 2179-83 (1990).

[51] *See generally* MICHAEL J. SANDEL, DEMOCRACY'S DISCONTENT: A NEW EDITION FOR OUR PERILOUS TIMES (2022) (arguing that democratic societies require a strong sense of community and active citizen participation, not merely individuals operating in a market-driven system).

[52] Charles Taylor, *Why Democracy Needs Patriotism*, *in* FOR LOVE OF COUNTRY: DEBATING THE LIMITS OF PATRIOTISM 119, 120 (Martha C. Nussbaum ed., 1996).

general."[53]

Moreover, public deliberation demands not only our energy, but shared understandings that make communication and rational persuasion possible. The kind of discursive preconditions that make rational discourse possible are created by our membership in a political community. The community's speech norms, which constrain our actions and words, create an environment in which rational discourse can take place.[54]

The norms governing public discourse, therefore, depend on shared understandings that allow effective communication among participants. Public discourse requires participants to make sacrifices, such as presenting their claims in a way that benefits the public good rather than simply defending their own private interests.[55] While this is a general requirement of deliberation, the specific ways in which these translations into the public good are enacted—that is, the very intelligibility of public debate—depend on the cultural understandings that are acquired through socialization in a particular community.

The very possibility of having our opinions influenced, and even reformed, in the light of mutual discussion, "necessarily implies a degree of cohesion. To some extent, the members must know one another, listen to one another, and understand one another. If they are not mutually acquainted, or if they cannot really understand one another, how can they truly engage in joint deliberation?"[56] If the speech norms that govern this process are violated, this "almost invariably makes speech seem . . . intimidating or abusive."[57]

If democracy is predicated upon the existence of a political community, and if that community is shaped and indeed

---

[53] *Id.*

[54] *See* Robert C. Post, *Community and the First Amendment*, 29 ARIZ. ST. L.J. 473, 482 (1997) ("[P]ublic discourse can link the individual wills of citizens to the general will of the nation only if public discourse is perceived as a process of rational dialogue, and community rules of civility are what define and constitute dialogue as rational.").

[55] *See generally* JOHN RAWLS, POLITICAL LIBERALISM (2005) (arguing that citizens should justify their political decisions with reasons accessible to all in public discourse, especially when addressing constitutional essentials and matters of basic justice).

[56] Charles Taylor, *The Dynamics of Democratic Exclusion*, 9 J. DEMOCRACY 129, 143 (1998).

[57] Post, *supra* note 54, at 482.

created by its members' discourse, then social-media platforms shoulder a vast responsibility in the preservation of a viable democratic community.

Social media, as we shall explore, disrupts two critical processes in communication that facilitate community cohesion. First, social media disrupts the shared experience of mass media consumption that allows everyone to participate in a collective conversation, thereby constituting what we know as "public opinion." Second, social media also undermines the ability of communities to institute and uphold common speech norms—norms that both define communities' identities and facilitate public debate. We delve into each of these impacts in the subsequent sections.

B. Media as the Infrastructure of the Public

In the democratic polity, the government is meant to respond to the currents of public opinion. Democratic legitimacy exists only when citizens have a warranted belief that they are able to shape their common political fate through engagement in public debate.[58] Since democracy is, to a large extent, "government by public opinion,"[59] the creation of a "public" is of foremost importance for the maintenance of democracy. Once a public is said to exist, we grant it agency. Publics are said to "scrutinize, ask, reject, opine, decide, judge, and so on."[60] Through this "extraordinary fiction,"[61] publics rule. They are said to "rise up, to speak, to reject false promises, to demand answers, to change sovereigns, to support troops, to give mandates for change, to be satisfied, to scrutinize public conduct, to take role models, to deride counterfeits."[62]

This construction of citizens as a sovereign "public" was only made possible by the rise of a new technological form of

---

[58] *See supra* note 54 and accompanying text.
[59] CARL SCHMITT, CONSTITUTIONAL THEORY 275 (Jeffrey Seitzer ed., 2008).
[60] MICHAEL WARNER, PUBLICS AND COUNTERPUBLICS 123 (2002).
[61] *Id.*
[62] *Id.*

communication: the various forms of mass media.[63] A "public opinion" is only sociologically possible where most individuals have access to common informational experiences.[64] This was noted since the emergence of newspapers by no less an observer than Alexis de Tocqueville: only a newspaper can claim the capacity to "succeed in putting the same thought in a thousand minds at the same instant."[65] The newspaper can mold "an immense, abstract, and sovereign crowd which it will name opinion."[66] The same is true for newer forms of mass media. According to media scholar Stig Hjarvard, the media serves as "a realm of shared experiences" by producing texts, videos, and sounds that are consumed by a large number of citizens.[67] In this way, sociologist Jeffrey Alexander argues, the media plays a central role in producing "the symbolic patterns that create the invisible tissues of society on the cultural level."[68] The public is distinguished from other forms of political communities by "the fact that its members ha[ve] access to the kind of publicness made possible by the printed word."[69]

By continuously disseminating content that engages a broad audience, mass-media outlets cultivate a public sphere in which consumers can imagine themselves both as individuals who perceive the world through multimedia platforms and as individuals who are perceived by the world. Whether through print, radio, or broadcast, consumers are presented with

---

[63] *See* Robert C. Post, *Data Privacy and Dignitary Privacy: Google Spain, the Right to Be Forgotten, and the Construction of the Public Sphere*, 67 DUKE L.J. 981, 1017-39 (2017).

[64] *See* JÜRGEN HABERMAS, THE STRUCTURAL TRANSFORMATION OF THE PUBLIC SPHERE: AN INQUIRY INTO A CATEGORY OF BOURGEOIS SOCIETY 181 (Thomas Burger trans., 1989); CHARLES TAYLOR, PHILOSOPHICAL ARGUMENTS 257-87 (1995).

[65] ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 906 (James Schleifer trans., 2012).

[66] GABRIEL TARDE, ON COMMUNICATION AND SOCIAL INFLUENCE 318 (Terry N. Clark ed., 1969).

[67] Stig Hjarvard, *The Mediatization of Society: A Theory of the Media as Agents of Social and Cultural Change*, 29 NORDICOM REV. 102, 126 (2008).

[68] Jeffrey C. Alexander, *The Mass-Media in Systemic, Historical, and Comparative Perspective*, *in* MASS MEDIA AND SOCIAL CHANGE 17, 18 (Elihu Katz & Tamás Szecskö, T. eds., 1981).

[69] JOHN B. THOMPSON, MEDIA AND MODERNITY: A SOCIAL THEORY OF THE MEDIA 126-27 (1995).

narratives, images, and information that shape their understanding of the world, their place in it, and their interactions with it. In segmented and multicultural societies, the media is a key social institution in fostering a sense of political community. Through its production of shared daily experiences and enforcement of speech norms, the media helps shape public discourse in ways that are directed toward society as a community.[70] Therefore, in effect, mass media creates "the conditions in which readers conceive themselves as a public, and hence as implicated in the formation of the public opinion that is essential to a democracy."[71]

These dynamics—the emergence of a "public" out of the creation of new communication technologies—can be generalized: the shape of the public is tied to a particular form of technological infrastructure.[72] Before the technical possibility of spreading the same message to a large and indeterminate number of people, there was no possibility for ordinary people to transcend the realm of their private lives and conversations.[73] As technology progressed, the relationship between individuals and the public sphere correspondingly changed. Internet search engines, for

---

[70] *See, e.g.*, JEFFREY C. ALEXANDER, THE CIVIL SPHERE 74 (2006) (arguing that public opinion "orients itself to the society qua collectivity, to an audience of citizens and to institutional actors only insofar as they are members thereof").

[71] Post, *supra* note 6363, at 1042.

[72] *See* Bernard Enjolras & Kari Steen-Johnsen, *The Digital Transformation of the Political Public Sphere: A Sociological Perspective*, *in* INSTITUTIONAL CHANGE IN THE PUBLIC SPHERE 99, 99 (Fredrik Engelstad et al. eds., 2017) ("[T]he media is conceived as the most important institutional communicative infrastructure of the public sphere."). *See also* Matthias Korn et al., *Infrastructuring Publics: A Research Perspective*, *in* INFRASTRUCTURING PUBLICS: THE NORDIC MODEL 11, 24 (Matthias Korn et al. eds., 2019) (describing how "science and technology studies, media and communication studies, socio-informatics, architecture and other disciplines have discovered the interwovenness of diverse infrastructures with publics as a major topic of interest").

[73] Post, *supra* note 6363, at 1017-39.

example, fundamentally changed this infrastructure by allowing access on demand to a virtually infinite number of content.[74]

Social-media platforms, through algorithmic recommendations and content moderation, are also fundamentally changing the way the public is constituted.[75] In the social-media world, people are offered customized versions of the news, which challenges the idea that we share the same normative and factual world.[76] An entity curating feeds in social media acquires a fundamental role in the formation of public opinion and therefore becomes a crucial feature of our current democracies.

C. Speech Norms as the Boundaries of the Public Sphere

A community is a type of social group in which individuals share certain features, such as family ties, geographical location, or common historical memory.[77] Belonging to a community comes to shape its members' identity:[78] a human

---

[74] *See* Post, *supra* note 63, at 990, 1045 (emphasizing how "Internet search engines underwrite the virtual communicative space in which democratic public opinion is now partially formed" and therefore "serve the same public interest as does the press").

[75] *See* Enjolras & Steen-Johnsen, *supra* note 72, at 102 ("Space-biased media in the form of digital communications entail a democratization of access to information and to the means of information production, undermining the cultural monopoly of knowledge and information that has been predominant with previous communication technology (print and electronic mass media). Such a transformation of the technological infrastructure of communication is not without far-reaching consequences in terms of the institutional structure and functions of the public sphere.").

[76] One of us has elaborated on this "digital epistemic divide." *See generally* Gilad Abiri & Johannes Buchheim, *Beyond True and False: Fake News and the Digital Epistemic Divide*, 29 MICH. TECH. L. REV. 59 (2022).

[77] MICHAEL J. SANDEL, LIBERALISM AND THE LIMITS OF JUSTICE 150 (1998) ("[C]ommunity describes not just what they have as fellow citizens but also what they are, not a relationship they choose [as in a voluntary association] but an attachment they discover, not merely an attribute but a constituent of their identity.").

[78] *Id.* In a widely accepted definition, Henri Tajfel defines social identity as "that part of an individual's self-concept, which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership." HENRI TAJFEL, HUMAN GROUPS AND SOCIAL CATEGORIES: STUDIES IN SOCIAL PSYCHOLOGY 255 (1981).

being becomes a self once she feels part of the Smith family, the Rapa Nui people, or the Nepalese nation.[79] As with every community, democratic publics require the enforcement of certain common norms, which place specific demands on their members and thus distinguish them from outsiders.[80]

The way *we* behave and speak is different from the way *they* behave and speak.[81] A Japanese child is expected to gradually recognize the different degrees of seniority in his community in ways that a visitor's child is not.[82] An Argentine man will greet other men with a kiss on the cheek in a way that would baffle a Brazilian.[83] As we see, while some norms are somehow present

---

[79] GEORGE HERBERT, MEAD, MIND, SELF, AND SOCIETY: FROM THE STANDPOINT OF A SOCIAL BEHAVIORIST 162 (Charles W. Morris ed., 1985) ("What goes to make up the organized self is the organization of the attitudes which are common to the group. A person is a personality because he belongs to a community, because he takes over the institutions of that community into his own conduct.").

[80] *See* JOSEPH R. GUSFIELD, COMMUNITY: A CRITICAL RESPONSE 29 (1978) ("[T]he crucial quality of communal interaction is the recognition that a common identity of communal membership implies special claims which members have on each other, as distinct from others.").

[81] Cliff Goddard & Anna Wierzbicka, *Discourse and Culture*, *in* 2 DISCOURSE AS SOCIAL INTERACTION 231, 231 (Teun A. Van Djik ed., 1997) ("In different societies people not only speak different languages and dialects, they use them in radically different ways.").

[82] *See* RUTH BENEDICT, THE CHRYSANTHEMUM AND THE SWORD: PATTERNS OF JAPANESE CULTURE 43-75 (2005) (describing Japan's complex social hierarchies and how social norms are shaped around it); *id.* at 47 (noting specifically that "[e]very greeting, every contact must indicate the kind and degree of social distance between men . . . There is a different 'you' that must be used in each case and the verbs have different stems"). *See also* Goddard & Wierzbicka, *supra* note 81, at 233 ("For example, among the Wolof of West Africa exchanging greetings [nuyyu or dyammanté] is a highly structured routine. A cultural outsider would never realize it, but each greeting exchange establishes the relative rankings of the participants.").

[83] *See* Alessandro Duranti, *Universal and Culture-Specific Properties of Greetings*, 7 J. LINGUISTIC ANTHROPOLOGY 63, 63 (1997) ("[D]espite the attention greetings have received in the social sciences, there is to date no generalizable definition of greetings and therefore no systematic way for deciding what qualifies as 'greetings' in a particular speech community.").

in every community (prohibition of murder, to give the most obvious example), the precise shape of community norms is arbitrary, as they need to distinguish one community from the other. For instance, not only do the specific ways of showing politeness vary from culture to culture, but the very meaning of politeness is extraordinarily context-dependent.[84]

The enforcement of community norms through punishment of violators helps define a community's limits, illustrating the degree and contours of personal liberty the community allows. Changes that challenge or weaken these norms can blur the boundaries of the community; if the norms disappear, so does the community.[85] This is true when it comes to a football fan club, a religious group, or the citizens of a country. For this reason, Richard Rorty defines an immoral act as "the sort of thing which, if done at all, is done only by animals, or by people of other families, tribes, or historical epochs. If done by one of us, or if done repeatedly by one of us, that person ceases to be one of us. She is an outcast . . . ."[86]

Speech norms are a vital part of the norms that hold communities together.[87] In the same way that there are some

---

[84] *See generally* Richard J. Watts, *Linguistic Politeness and Politic Verbal Behaviour: Reconsidering Claims for Universality*, *in* POLITENESS IN LANGUAGE: STUDIES IN ITS HISTORY, THEORY AND PRACTICE 43, 43 (Richard J. Watts, Sachiko Ide & Konrad Ehlich eds., 2019) ("We cannot be at all certain that an English native Speaker . . . understands 'politeness' in exactly the same way as the German native Speaker understands 'Höflichkeit' or the French native Speaker 'politesse'.").

[85] We follow Robert Post, who defines a community "as a social formation that inculcates norms into the very identities of its members. So far from being considered autonomous, persons within a community are understood to depend, for the very integrity and dignity of their personalities, upon the observance of these norms." Post, *supra* note 18, at 286.

[86] RICHARD RORTY, CONTINGENCY, IRONY, AND SOLIDARITY 59-60 (1989); *see also* KAI T. ERIKSON, WAYWARD PURITANS: A STUDY IN THE SOCIOLOGY OF DEVIANCE 11 (1966) ("The deviant is a person whose activities have moved outside the margins of the group, and when the community calls him to account for that vagrancy it is making a statement about the nature and placement of its boundaries. It is declaring how much variability and diversity can be tolerated within the group before it begins to lose its distinctive shape, its unique identity.").

[87] Technology makes possible to imagine a case in which different people inhabiting the same geographical space develop radically different speech norms to talk to one another. In this world, people do not communicate with

things that should (or should not) be *done* in society (taking one's shoes off in the mosque, or taking care of elderly parents) there are other things that should or should not be *said*, or said in certain ways. The law draws on these shared understandings to regulate conduct in a way that resonates with the addressees of legal rules, and in turn, it reinforces the social norms from which it takes inspiration.[88] Let us consider three major examples concerning speech norms: insults, hate speech, and misinformation.

Insults are a clearcut example of an extremely context-dependent community norm. What constitutes an insult, the gravity of being insulted in public, or expectations of legal remedy after being insulted vary greatly from society to society: a reference to one's skin color can be a brutal insult in some

---

their physical neighbors as much as they do across different channels such as specific mass media or in specific online fora. In an extreme case, one half of the population adheres to a set of community norms enforced by their own authorities, while the other half adheres to a different set of norms and a different set of enforcers. In such a case, we would need to talk about two communities and not about one; and finding ways to make these communities coexist in a same geographical space would be a distinct political problem. Robert Post has famously argued that the highly idiosyncratic character of free speech law in the United States responds to the country's fragmentation in many communities with different community norms among which the Supreme Court would refuse to pick. *See* Robert C. Post, *The Constitutional Concept of Public Discourse: Outrageous Opinion, Democratic Deliberation, and Hustler Magazine v. Falwell*, 103 HARV. L. REV. 601, 634-35 (1990). To the extent that extreme polarization in the contemporary United States can be said to bring about two different communities with distinct speech norms, our analysis would hardly be applicable to the United States. On the other side, it is still likely that even in this context some speech norms apply widely in the whole territory of the United States. However, the extent to which these statements are true is an empirical question beyond the reach of this Article.

[88] This is a type of "double institutionalization," which is an idea developed by legal anthropologist Paul Bohannan which refers to the idea that law is both embedded in social institutions and also exists as a separate and distinct institution. In other words, law is both created by social institutions and then re-created by legal institutions. *See generally* Paul Bohannan, *The Differing Realms of the Law*, 67 AM. ANTHROPOLOGIST 33 (1965).

countries and a rather innocent joke in others;[89] use of certain pronouns to address someone from a distinct social position can be an inexorable demand in some places and an incomprehensible one in others.[90] Tellingly, insults are famously among the words that are more difficult to translate.[91] Law, even in modern democracies, draws heavily on these understandings to punish deviant behavior: while U.S. courts would demand an extremely high level of aggressiveness to sanction someone for insulting their neighbor, courts in Germany routinely accept such claims.[92]

Hate speech is a more complex version of the same phenomenon. The type of expressions that are commonly dubbed "hate speech" are impossible to define and identify without taking into account the particularities of a given society—its history, its different ways of describing itself, its

---

[89] Cross-cultural communication makes this type of incident ever more frequent. For instance, on an Instagram post, Manchester United Uruguayan player Edinson Cavani thanked a Uruguayan fan in Spanish: "Gracias Negrito"—an expression meant to be an affectionate one, which would be literally translated as "thank you, little Black one." Upon a warning about how the phrase might be interpreted in the United Kingdom, Cavani apologized, deleted the post, and was sanctioned by the club. *See* Jamie Jackson, *Edinson Cavani Apologises for Using Racial Term in Instagram Post*, GUARDIAN (Nov. 30, 2020), https://www.theguardian.com/football/2020/nov/30/manchester-united-edinson-cavani-apologies-for-racist-instagram-post [https://perma.cc/SN22-HAVR]. However, back in Uruguay, the case ignited a fierce reaction since the term "negrito" is not used as a slur. The Uruguayan Academy of Letters, for instance, published a harsh open letter, accusing the English Football League of "ignorance" and "poverty of cultural and linguistic knowledge." *See La Academia Uruguaya de Letras Rechazó la Sanción de la Federación Inglesa Contra Cavani*, INFOBAE (Jan. 2, 2021), https://www.infobae.com/america/deportes/futbol-europeo/2021/01/02/la-academia-uruguaya-de-letras-rechazo-la-sancion-de-la-federacion-inglesa-contra-cavani [https://perma.cc/7D8E-EW26].

[90] *See* James Q. Whitman, *Enforcing Civility and Respect: Three Societies*, 109 YALE L.J. 1279, 1295 (1999) (identifying diverging social customs in France, Germany, and the United States regarding "[e]veryday rituals of linguistic respect" such as forms of address).

[91] Pedro J. Chamizo-Domínguez, *Problems Translating Tabooed Words from Source to Target Language*, *in* OXFORD HANDBOOK OF TABOO WORDS AND LANGUAGE 199, 199 (Keith Allan ed., 2018) ("If the translator's job is always difficult, translating tabooed words is particularly problematic").

[92] *See* Whitman, *supra* note 90, at 1295-1313.

relations between majorities and minorities.[93] A phrase can be an unnoticed transgression, or even completely innocent, in some places, and profoundly hateful in others. To be able to distinguish between acceptable expression of disagreement on issues of race, religion, and gender, and hateful speech, we have to refer to community norms. Only an individual socialized "in any culture can tell whether any given communication is 'extreme,' meaning that the communication violates essential standards of civility and hence is vulnerable to legal sanction."[94]

Misinformation is likewise identified and defined through a given community's epistemic norms. Deployment of these norms is not a way of establishing scientific truth, but rather a way to engage in the "social construction of reality."[95] In the same way that parents instill in their children the correct way to behave, they impart to them an understanding of the acceptable ways to gain knowledge about the world. When a kid tells a parent that a classmate told them that the moon is made out of cheese, they will tell them to look it up on Wikipedia or ask their teacher. For a more dramatic example, take the infamous imprisonment of Galileo Galilei for saying the Earth revolves around the Sun, against the established knowledge inherited from the Biblical tradition. The Church's reaction to Galileo was not a way of defending truth, but rather a defensive reaction to the threat the scientific method represented to their religion-based epistemic norms, and therefore on the existence of the society dominated by the

---

[93] *See* Robert Post, *Hate Speech*, *in* EXTREME SPEECH AND DEMOCRACY 123, 125 (Ivan Hare & James Weinstein eds., 2009) ("All legal attempts to suppress hatred, whether of racial groups or of the King, must face a profound conceptual difficulty. They must distinguish hatred from ordinary dislike or disagreement . . . [which] is the lifeblood of politics.").

[94] *Id.* at 128.

[95] *See generally* PETER L. BERGER & THOMAS LUCKMANN, THE SOCIAL CONSTRUCTION OF REALITY: A TREATISE IN THE SOCIOLOGY OF KNOWLEDGE 14 (1990) (arguing that "reality is socially constructed and that the sociology of knowledge must analyse the process in which this occurs").

Church.[96] In fact, the victory of the scientific approach to reality would change the world a few decades later.

These examples, simple in appearance, show a complex structure. Distinguishing appropriate teasing from an offensive insult demands high competence in a specific culture.[97] Social and legal sanctions for expressions that go beyond acceptable aggression show people the appropriate level of respect they can demand from society and thereby contribute to social cohesion. While these dynamics can be straightforward and unproblematic when it comes to small, premodern communities, they become more contentious in liberal, diverse democracies. We explore liberal, democratic communities next.

III. Böckenförde's Paradox

Democratic communities stand in a fundamental paradox: they require common speech norms to prosper and grow, but must avoid overly stringent enforcement to uphold their liberal and democratic ethos. The fundamental structure of this paradox was most prominently spotted by German constitutional theorist Ernst Böckenförde. He coined what has since become known as the Böckenförde paradox: "[t]he liberal, secularized state is sustained by conditions it cannot itself guarantee."[98] Accordingly, if a liberal state "used its own instruments, namely legal coercion and authoritative command [to enforce communitarian values,] it would lose its liberal nature and relapse into the old totalitarian claim."[99] Free

---

[96] *See* David C. Lindberg, *Galileo, the Church, and the Cosmos*, *in* WHEN SCIENCE & CHRISTIANITY MEET 33, 47-48 (2003) ("The larger issue that lay behind this question was that of epistemological authority: are cosmological truth-claims dependent on science or on theology—on conclusions drawn from reason and sense experience or on the content of biblical revelation as interpreted by the fathers of the church?").

[97] Post, *supra* note 93, at 128 ("It is by reference to norms that a well-socialized person in any culture can tell whether any given communication is 'extreme,' meaning that the communication violates essential standards of civility and hence is vulnerable to legal sanction.").

[98] Ernst-Wolfang Böckenförde, *The Rise of the State as a Process of Secularization*, *in* RELIGION, LAW, AND DEMOCRACY: SELECTED WRITINGS 152, 167 (Mirjam Künkler & Tine Stein eds., 2020).

[99] Tine Stein, *The Böckenförde Dictum—On the Topicality of a Liberal Formula*, 7 OXF. J.L. & REL. 97, 101 (2018).

speech scholars have noted a similar structural feature of public discourse, which "can sustain democratic legitimation only if it is conducted with a modicum of civility, yet the enforcement of civility constrains freedom of speech."[100] In democracies, therefore, the civility norms that maintain community together must and must not be upheld by the state.

The classical example of state-sponsored enforcement of social norms is state religion. The maintenance of a state religion can foster, and arguably has fostered, higher degrees of social cohesion.[101] During most of human history, the state had a vested interest in ensuring that its subjects shared religious ties and complied with religious rules, which in turn legitimated and strengthened state rules and commands.[102] Indeed, many advanced democracies today still maintain some institutional connection with majoritarian religions.[103] However, as democracies become more secular and liberal, these ties to religion become problematized and weakened. Some degree of separation between state, individual rights, and religion is often regarded as one of the markers of a truly liberal state.[104]

---

[100] Post, *supra* note 63, at 1009.

[101] For an empirical study thereto, see Annette Schnabel & Florian Groetsch, *In God We Trust – The Role of Religion for Cohesion and Integration in Current European Societies*, 1 EUR. J. CULTURAL & POL. SOCIO. 375 (2014).

[102] Paul Dumouchel, *Politics and the Permanence of the Sacred*, *in* HANDBOOK OF POLITICAL ANTHROPOLOGY 115, 115 (Harald Wydra & Bjørn Thomassen eds. 2018) ("During most of human history (and still today) politics and the sacred would have shared a stage.").

[103] *See e.g.*, Nahshon Perez, *Hegemonic Religions, Majoritarianism, and the Legitimate Limits of Governmental Religious Bias*, 11 RELIGIONS 438, 439-40 (2020) (defining this "Religious-Majoritarian-Approach" as a "well-known approach that is embodied in various European countries" that has also "gained wide legal recognition from leading courts, such as the European Court of Human Rights, and received scholarly justification from important political and legal theorists").

[104] *Id.* at 445 ("Now, many liberal theorists, as well as advocates of the model of separation between religion and state, would reject any attempt by a

While the presence of religion in the public sphere has decreased, it remains true that any democratic state demands a sense of common identity which leads to an inner motivation to comply with rules.[105] However, it is impossible for the state to enforce the community norms that are required to maintain a shared sense of belonging beyond a certain point—at least not without ceasing to be a liberal state. This is a distinctively liberal challenge: illiberal states, such as theocracies or totalitarian dictatorships, do not run into this problem, as they have no internal barrier against state enforcement of ideological tenets.

Take a communicative activity that acts as a social cohesive: gossip. Gossip is so central to human interaction that anthropologist Robin Dunbar even conjectured that humans created language for its sake.[106] In small communities, gossip provides valuable information about other people's character and trustworthiness.[107] In larger communities, gossip about celebrities or otherwise notable individuals provides a common informational experience that helps strangers interact meaningfully.[108] Being such an important activity, gossip is severely regulated by social norms, which normally establish a

---

democratic government to adopt the [Religious-Majoritarian-Approach] in principle. According to their view, a liberal state should never side with a given religion . . . .").

[105] *See* DAVID MILLER, ON NATIONALITY 90-99 (1995) (generally arguing that "states . . . are likely to function most effectively when they embrace just a single national community").

[106] This is the general thesis in ROBIN DUNBAR, GROOMING, GOSSIP, AND THE EVOLUTION OF LANGUAGE (1996).

[107] *See* Charlotte J. S. De Backer & Maryanne L. Fisher, *Tabloids as Windows into Our Interpersonal Relationships: A Content Analysis of Mass Media Gossip from an Evolutionary Perspective*, 6 J. SOC., EVOLUTIONARY & CULTURAL PSYCH. 404, 408 (2012) (providing a list of "multiple functions of gossip" in the literature including the "promot[ion of] within-group solidarity" and being a "device to learn group norms, values, and one's place within a group").

[108] *See* Frank McAndrew, *Gossip is a Social Skill – Not a Character Flaw*, CONVERSATION (Jan. 20, 2016), https://theconversation.com/gossip-is-a-social-skill-not-a-character-flaw-51629 [https://perma.cc/L4W9-KQQC] ("Because of the familiarity we feel with celebrities, they can serve an important social function: they may be the only "friends" we have in common with new neighbors and coworkers. They're shared cultural touchstones that facilitate the types of informal interactions that help people become comfortable in new surroundings.").

complex equation that combines, among other factors, how close one is to the person one is gossiping with and about, the gravity of one's accusation, and the trustworthiness of the source of information.[109] And yet, we would be aghast if a liberal government wanted to enforce these norms legally onto individuals, which would constitute a violent invasion of privacy and free speech. Indeed, when the news arrived that a city in the Philippines outright banned gossip, Western media were quick to dismiss the measure as "draconian."[110]

All of that said, as is evident from their continued existence, constitutional liberal states have somehow overcome the Böckenförde paradox—in practice, if not in theory. They accomplished this through three main strategies: (1) socialization in state-directed institutions, (2) direct regulation of speech, and (3) outsourcing to civil society institutions.

(1) The first strategy is the indirect inculcation of community norms through institutions, such as public schools, the military, and the family. These are settings in which liberal societies are generally comfortable with enforcing speech and civility norms, often in ways that would seem at odds with democratic principles. Public schools' ability to punish students for hateful or rude speech, for example, makes them a powerful tool for spreading common community norms. Notably, these

---

[109] *See generally* De Backer & Fisher, *supra* note 107 (describing different rules at play in the exercise of gossip).

[110] *See e.g.*, Harry Howard, *The Town that Banned GOSSIP: People Who Spread Rumours are Fined $10 and Forced to Pick up Litter in Philippine Neighbourhood*, DAILY MAIL (Apr. 26, 2019), https://www.dailymail.co.uk/news/article-6964661/People-spread-rumours-fined-10-forced-pick-litter-Philippine-town.html [https://perma.cc/FM5S-68QF] ("A town in the Philippines has taken the draconian step to ban gossip."); Hannah Ellis-Petersen, *Have You Heard About the Philippine Mayor Who Banned Gossip?*, GUARDIAN (May 1, 2019), https://www.theguardian.com/world/2019/may/01/have-you-heard-about-the-philippine-mayor-who-banned-gossip [https://perma.cc/6F5L-SGTM] ("Across southeast Asia, the introduction of laws against fake news has become increasingly common, but has also been linked to a suppression of freedom of speech.").

institutions are seen as outside the purview of regular free expression protections even in the United States.[111] They are sites in liberal societies where community norms can be promoted and enforced.

However, even in these institutions, the state is limited in the extent to which it can use its institutions to enforce community norms without ceasing to be liberal. States do usually have a reasonably strong grip on the school system—but kids grow up and leave. States can organize military parades and patriotic festivals—but people can choose not to attend.[112] States can tell a family that they will take their child if they fail to educate them in a certain way—but they will realistically do so in extremely exceptional cases.[113] Even if the state decides to use all of its symbolic power over these institutions, it still needs community norms to be enforced in the innumerable other realms that constitute civic life.

(2) The second strategy at the state's disposal is the direct regulation of speech. As an illustration, under the common law of libel, which held sway in the United States prior to its

---

[111] *See generally* Robert C. Post, *There is No 1st Amendment Right to Speak on a College Campus*, VOX (Dec. 31, 2017), https://www.vox.com/the-big-idea/2017/10/25/16526442/first-amendment-college-campuses-milo-spencer-protests [https://perma.cc/R74A-QSAX] (arguing that "members of the university . . . have the right to *academic freedom*, not First Amendment freedom of speech. Academic freedom is defined in terms of the twin missions of the university; it encompasses freedom of research and freedom of teaching."). The United States Supreme Court, arguably the most protective of free speech worldwide, routinely recognizes limitations on students' free speech based on schools' educational mission. *See* Bethel School District No. 403 v. Fraser, 478 U.S. 675, 685 (1986) ("The First Amendment does not prevent the school officials from determining that to permit a vulgar and lewd speech . . . would undermine the school's basic educational mission.").

[112] *See generally* Marie-Christin Gabriel et al., *Embodying the Nation: The Production of Sameness and Difference in National-Day Parades*, 21 ETHNOGRAPHY 506 (2020) (describing the significance of parades for the construction of national identity).

[113] Eur. Ct. Hum. Rts. [GC], Strand Lobben and Others v. Norway (case 37283/13, Sept. 10, 2019) (judging that measures that have "the consequence that the applicants' legal ties with the child are definitively severed . . . should only be applied in exceptional circumstances and could only be justified if they were motivated by an overriding requirement pertaining to the child's best interests") (internal quotations omitted).

restriction in *New York Times Co. v. Sullivan*,[114] an individual could receive compensation for a false and defamatory statement. To succeed in such a case, it was sufficient for the claimant to demonstrate that a presumably false statement that was made public was defamatory in nature and referred to the claimant. There was no requirement for the claimant to demonstrate *actual* harm to their reputation.[115] Academics have persuasively maintained that the expectation of damages is an aspect of "enforcement of society's interest in its rules of civility."[116] By paying attention to the objective violation of community norms per se, and not to actual psychic or reputational damage, the state is drawing on community norms to directly mandate what it allows people to say in public.

This is just one instance. In fact, even in the domain of public discourse, the notion that legislation can ban vulgarity and other types of speech considered contrary to community speech norms has a rich history in common law, filled with caselaw about group libel,[117] fighting words,[118] true threats,[119] words that might incite imminent lawlessness,[120] public nudity,[121] broadcast indecency,[122] obscenity,[123] and child pornography.[124] As soon as one looks closely, it becomes clear

---

[114] 376 U.S. 254 (1964).

[115] *See generally* Philip A. Hamburger, *The Development of the Law of Seditious Libel and the Control of the Press*, 37 STAN. L. REV. 661 (1985) (laying out the history and development of the law of seditious libel).

[116] Robert C. Post, *The Social Foundations of Defamation Law: Reputation and the Constitution*, 74 CALIF. L. REV. 691, 711 (1986).

[117] *See, e.g.*, Beauharnais v. Illinois, 343 U.S. 250 (1952).

[118] *See, e.g.*, Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).

[119] *See, e.g.*, Virginia v. Black, 538 U.S. 343 (2003).

[120] *See, e.g.*, Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (holding that speech advocating lawless action is protected unless it is "directed to inciting or producing imminent lawless action and is likely to incite or produce such action").

[121] *See, e.g.*, City of Erie v. Pap's A.M., 529 U.S. 277 (2000); Barnes v. Glen Theatre, Inc., 501 U.S. 560 (1991).

[122] *See, e.g.*, FCC v. Pacifica Found., 438 U.S. 726 (1978).

[123] *See, e.g.*, Miller v. California, 413 U.S. 15 (1973).

[124] *See, e.g.*, New York v. Ferber, 458 U.S. 747 (1982).

that the decisionmaker can assess whether there was a transgression only insofar as she contrasts the utterance with some standard, which is socially determined–a community norm.[125]

Direct speech regulation, however, is a limited technique when it comes to enforcing community norms. Due to its limited capacity, the criminal system must select which transgression it will enforce in practice.[126] This is particularly salient when it comes to the enforcement of speech norms, as speech is ubiquitous. Even if it was technologically and bureaucratically possible for the state to control every single utterance in society to ensure it conforms to civility norms, this sole prospect is the raw material for classical dystopian novels.[127] Even when they engage in legal enforcement of civility norms, states usually delegate this power to individuals, making the course of both criminal and civil action depend on whether some (very) affected individual wishes to go through the difficult task of utilizing state law against her aggressor.[128]

---

[125] This is frequent whenever the law imposes "reasonableness" or similar standards to assess conduct: standards such as the "reasonable person" are given content only through the conscious and unconscious social deliberation that leads up to "the social norms of the particular community." In this way, "cultural norms truly become the law." Paul T. Hayden, *Cultural Norms as Law: Tort Law's "Reasonable Person" Standard of Care*, 15 J. AM. CULTURE 45, 53 (1992). Robert Post, for instance, has frequently noted this phenomenon in the context of norms regulating speech. *See e.g.*, Post *supra* note 93, at 128 ("The law commonly enforces social norms of this kind, as for example when it prohibits defamation, invasions of privacy, intentional infliction of emotional distress, flag burning, and so on.").

[126] William J. Stuntz, *The Pathological Politics of Criminal Law*, 100 MICH. L. REV. 505, 506-07 (2001) ("Anyone who reads criminal codes in search of a picture of what conduct leads to a prison term, or who reads sentencing rules in order to discover how severely different sorts of crimes are punished, will be seriously misled. The reason is that American criminal law . . . is very broad; it covers far more conduct than any jurisdiction could possibly punish.").

[127] This was precisely the role of the telescreens in the classical dystopian novel *Nineteen Eighty-Four*. GEORGE ORWELL, NINETEEN EIGHTY-FOUR (1949).

[128] *See e.g.*, Scott Griffen, *Defamation and Insult Laws in the OSCE Region: A Comparative Study*, ORG. FOR SEC. & CO-OPERATION EURO. 10 (Mar. 2017), https://www.osce.org/files/f/documents/b/8/303181.pdf

This self-restraint helps keep state power limited, and therefore allows for a degree of freedom for community norms to evolve organically,[129] but does also leave a huge space in which civility norms remain uncontrolled, at least by the state.

(3) Liberal states possess the tools to maintain community but refuse to fully use these abilities to remain liberal, and as Böckenförde taught us, the force to keep society together cannot come solely from the state. [130] Thus, if a state wants to protect its underlying community it has the difficult task of delegating the enforcement of community-maintaining onto others. The institutions that usually step into the gap left by the absence of the state are the so-called civil-society organizations: public-facing institutions that are separated from both the state and the market, such as educational institutions, mass-media, non-governmental organizations, neighborhood associations, unions, affinity groups, and religious institutions.[131]

Social theorists of different persuasions have long acknowledged the role of civil society in maintaining the collective identity that makes governance possible, "drag[ging]

[https://perma.cc/8NVA-ZFBM] ("The system of private prosecution is fairly common for defamation cases in the [Organization for Security and Co-operation in Europe] region, and most states at least require that charges can only be brought upon the request of the party claiming to have been defamed.").

[129] *See e.g.*, Post *supra* note 93, at 129-30 ("[Social] norms are shared and yet evolve over time. . . . [P]recisely because norms evolve, they are intrinsically contestable. There are constant struggles over the developing meaning of shared standards and expectations. . . . [The law] must continuously choose what kind of community it will sustain. It must always decide whether to reinforce existing norms or to allow for the growth of new norms.").

[130] Böckenförde *supra* note 98, at 167 ("[A] liberal state . . . can only survive if the freedom it grants to its citizens is regulated from within, out of the moral substance of the individual and the homogeneity of society.").

[131] In a classic definition going back to Georg Hegel and Antonio Gramsci, civil society is conceptualized as "the realm of political, cultural, legal, and public life that occupie[s] an intermediate zone between economic relations and political power." *See* ALEXANDER, *supra* note 70, at 28.

people into the general torrent of social life."[132] According to sociologist Manuel Castells, for example, civil society is "a set of organizations and institutions . . . which reproduce, albeit sometimes in a conflictive manner, the identity that rationalizes the sources of structural domination."[133] For sociologist Jeffrey Alexander, "civil society should be conceived as a solidary sphere, in which a certain kind of universalizing community comes to be culturally defined and to some degree institutionally enforced."[134] This role of civil society institutions in enforcing the political identity that enables the state to govern is indeed so strong that in a classical Marxist account, civil society institutions are "ideological apparatuses" of the state itself, and enforcement of norms by civil society is, in fact, state enforcement of norms.[135]

Civil society fosters and enforces community norms in various ways. The majority religion usually mirrors and legitimates hierarchies in broader society.[136] Universities promulgate norms by which many people outside the university tell truth and falsehood apart.[137] Parent-teacher

---

[132] EMILE DURKHEIM, THE DIVISION OF LABOR IN SOCIETY 28 (Free Press, 1964) ("A nation can be maintained only if, between the State and the individual, there is intercalated a whole series of secondary groups near enough to the individuals to attract them strongly in their sphere of action and drag them . . . into the general torrent of social life.").

[133] Manuel Castells, *Communal Heavens: Identity and Meaning in the Network Society*, *in* THE POWER OF IDENTITY 1, 8 (2011).

[134] ALEXANDER, *supra* note 70, at 31.

[135] Louis Althusser, *Ideology and Ideological State Apparatuses*, *in* ON THE REPRODUCTION OF CAPITALISM: IDEOLOGY AND IDEOLOGICAL STATE APPARATUSES 232, 232 (G.M. Goshgarian trans., 2014).

[136] *See e.g.*, D.C. Leege, *Religion and Politics: United States*, *in* INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL & BEHAVIORAL SCIENCES 13040, 13040-41 (Neil J. Smelser & Paul B. Baltes eds., 2001) (explaining that religion helps to establish moral order for a society, legitimating social hierarchies, collective goals, and cultural boundaries). *See also* ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 66 (2000) ("Churches provide an important incubator for civic skills, civic norms, community interests, and civic recruitment.").

[137] ROBERT C. POST, DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM: A FIRST AMENDMENT JURISPRUDENCE FOR THE MODERN STATE 68, 76 (2012) (explaining that "[u]niversities and university faculty are the unique and primary sites in modern American society for the creation and diffusion

associations instill a sense of civic duty and shared bonds between parents, students, and educators,[138] as do voluntary citizens' organizations.[139] However, delegating this role comes with trade-offs. These civil-society institutions may enforce norms in ways that lack transparency, due process, and democratic accountability, potentially leading to biased or discriminatory outcomes. Arguably, however, the most important civil-society institution structuring public debate and enforcing community norms regarding speech in modern societies is the mass media. We turn to it now.

IV. Civil-Society Gatekeepers and the Bypass Effect

The rise of digital media fundamentally challenged civil society's control of public debate. Social media allows people to reach mass audiences without relying on the publication channels dominated by mass media. This "bypass," enabled by social media, undermines the ability of civil society "gatekeepers" to enforce community speech norms and structure public debate. While scholars have paid attention to aspects of this feature, we posit that one crucial question has not yet been explored enough: how this fundamental change in communication affects not only public debate, but the very stability of the democratic political community.

---

of disciplinary knowledge in the service of the public good" and that, therefore, "[t]hrough their virtual monopoly of graduate training, universities certify experts, as well as their expertise.").

[138] *See* PUTNAM *supra* note 136, at 290 ("[B]elonging to [a parent-teacher association] almost certainly inculcates civic skills in parents. . . . What's more, [a parent-teacher association] serves to establish and enforce norms of commitment and performance on the part of school officials, teachers, and perhaps even students.").

[139] *See e.g.*, John A. Booth & Patricia Bayer Richard, *Civil Society, Political Capital, and Democratization in Central America*, 60 J. POL. 780, 780-81 (1998) (stressing the importance of voluntary citizens' associations for the development of political behaviors and attitudes and citing literature).

A. Mass Media

Among civil-society institutions, mass media has a privileged relationship with the community norms regulating speech. Mass-media elites are the gatekeepers and curators of what can be said in public discourse without meeting social sanction.[140] They play an important role in shaping public debate and ensuring that it adheres to commonly accepted standards of civility.[141] Take for example the United States before digitalization: while neo-nazi literature was protected under First Amendment doctrine, civility norms followed and enforced by publishers made it nearly impossible to get it published with any reputed press.[142]

By choosing which content and forms to publish or broadcast, or by emphasizing and problematizing certain expressions, mass media gives a very clear sense of what kind

---

[140] *See, e.g.*, PAMELA J. SHOEMAKER & TIMOTHY VOS, GATEKEEPING THEORY 1 (2009) (defining "gatekeeping" as the "process of culling and crafting countless bits of information into the limited number of messages that reach people each day. . . . People rely on mediators to transform information about billions of events into a manageable subset of media messages"). *See also id.* at 3 ("The mass media provide a conduit through which new ideas and the groups who introduce them can reach the larger public. Media exposure is a contingent condition for acceptance of new ideas, and groups fight for access to the media and, therefore, to the audience.").

[141] *See* Abiri *supra* note 5, at 764 ("Before the advent of digital platforms, mass media had control of the bottleneck of information: without their publication capacity, it was nearly impossible to reach wide audiences. Mass media elites thus were able to exclude and sanction uncivil speech, such as hate speech . . . ").

[142] *See* Helen Young & Geoff M Boucher, *How Self-publishing, Social Media and Algorithms Are Aiding Far-right Novelists*, CONVERSATION (May 29, 2022), https://theconversation.com/how-self-publishing-social-media-and-algorithms-are-aiding-far-right-novelists-183453 [https://perma.cc/5NJB-2AVS] ("Publishing houses once refused to print such books, but changes in technology have made traditional publishers less important.").

of speech is acceptable "in polite company":[143] swearing,[144] taboos,[145] regional accents,[146] forms of dressing and makeup,[147]

---

[143] In the oral argument of *FCC v. Fox Television Stations, Inc.*, 556 U.S. 502 (2009), these words used in a dialogue between Fox's attorney and Justice Scalia nicely illustrated the role of mass media gatekeepers in enforcing civility norms. The FCC had sanctioned Fox because of the use of casual expletives during the Emmy Awards live broadcasting. Fox's attorney tried to argue that society had become less offended by this type of words:

> Carter Phillips: I believe that society is significantly more tolerant of these words today than it was 30 years ago.
>
> Antonin Scalia: Do you think your clients have had anything to do with that? [Laughter]
>
> Carter Phillips: In the scheme of things, probably very, very little to do with that compared to the way the language is used. Go to a baseball game, Justice Scalia. You hear these words every time you go to a ballgame.
>
> Antonin Scalia: You do, indeed, but you don't have them presented as something that is normal in polite company, which is what happens when it comes out in television shows.

See Transcript of Oral Argument FCC v. Fox Television Stations, Inc. (07-582), OYEZ, https://www.oyez.org/cases/2008/07-582 [https://perma.cc/TRY9-75NB] (cleaned up).

[144] *See id.*; *see also* David Marsh, *Is There Too Much Swearing in the Guardian?*, GUARDIAN (Apr. 14, 2010), https://www.theguardian.com/media/mind-your-language/2010/apr/14/swearing-guardian [https://perma.cc/47DH-YQQA] (discussing the newspaper's policy about swearwords).

[145] Cagri Yalkin & Ekant Veer, *Taboo on TV: Gender, Religion, and Sexual Taboos in Transnationally Marketed Turkish Soap Operas*, 34 J. MKTG. MGMT. 1149, 1149 (2018) (arguing that in some contexts "talking about taboo topics seen in soap operas enables consumers to speak about what they expect gender and religious norms to be").

[146] Jacqueline Kavanagh, *The BBC's Written Archives as a Source for Media History*, 5 MEDIA HIST. 81, 82 (1999) ("Discussions [among BBC producers as found in old reports] cover the whole area of presentation through speech—regional accents, pronunciation, pitch and tone of voice.").

[147] *See* Teri Finneman & Joy Jenkins, *Sexism on the Set: Gendered Expectations of TV Broadcasters in a Social Media World*, 62 J. BROAD. & ELEC. MEDIA 479, 487-89 (2018) (collecting TV anchors' reports about how they were reprimanded by not complying with traditional gender norms regarding dressing and hairstyle).

trustworthiness of information,[148] hate speech,[149] and an infinite et cetera.[150] As this enumeration suggests, mass media plays an overall conservative role in enforcing existing social norms. However, in societies as large and complex as ours, in the same way mass media gives us a "realm of shared experience,"[151] they also give us the shared rules for communicating civilly.

Therefore, in a media ecosystem where the means of communication were limited, traditional news media served as a key tool for civil society to promote and enforce community norms. To reach a wide audience, individuals or groups were dependent on the gatekeepers who controlled access to ears and eyeballs. This led to a situation where if media actors had a shared understanding of the boundaries of acceptable public speech, they could easily exclude speech and speakers that did not align with these standards from public discourse.

The gatekeeping powers of civil society generally, and mass media specifically, were crucial for their ability to keep the "public" together by enforcing a common set of speech norms.[152]

This ability was dependent on the specific technological ecosystem of mass media. For most of the twentieth century, the structure of information power was "asymmetrical and unidirectional."[153] The ability to print or broadcast information was limited and highly centralized. The information that was disseminated to the public was carefully curated and required significant financial resources to produce. This led to a concentration of power among a small number of organizations, such as major news networks, that had the capital to invest in broadcasting technology and the centralized

---

[148] *See supra* note 7676 and accompanying text.

[149] *See* Abiri *supra* note 5, at 761 ("Before digitalization, traditional media gatekeepers were one of the main social forces limiting hate speech.").

[150] Adrienne Chung & Rajiv N. Rimal, *Social Norms: A Review*, 4 REV. COMM. RES. 1, 18 (2016) ("[I]t is possible that greater immersion in the media environment in general . . . can also affect the perceived acceptability of a behavior" such as binge drinking or smoking. "Thus, exposure to media can influence normative perceptions, which in turn can affect behaviors.").

[151] Hjarvard, *supra* note 67, at 126.

[152] *See supra* Section II.C.

[153] Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 9 (2004).

production of information.[154] Mass media, therefore, created a "technological bottleneck, and the people who control mass media [were] gatekeepers controlling its use."[155]

Before digital media, this bottleneck enabled mass-media actors to be the "main institutional vehicle for regulating and enabling expression within society."[156] With limited means of communication, the role of mass media as gatekeepers of public discourse was crucial. The gatekeepers, such as journalists, editors, and shareholders, who controlled the production processes in print and broadcast media, had significant power in determining which topics were discussed on a large scale and what content was made available to the audience.[157]

B. The Bypass Effect

With the advent of digitalization, the role of traditional media has gradually shifted from "gatekeepers" to "gatewatchers."[158] Unlike centralized mass media, in which content is administered by a small number of entities, Internet-based media is highly decentralized, offering more extensive opportunities for information production and distribution, which become not only widely available but also incredibly cheap. This transformation "can be boiled down to one idea: it is no longer speech itself that is scarce, but the attention of

---

[154] *See* Enjolras & Steen-Johnsen, *supra* note 72, at 102 (emphasizing that "mass media communication required centralized means of information production and large investments in physical capital . . . ").

[155] Balkin, *supra* note 153, at 10.

[156] Enjolras & Steen-Johnsen, *supra* note 72, at 104-05.

[157] AXEL BRUNS, GATEWATCHING: COLLABORATIVE ONLINE NEWS PRODUCTION 11 (2005) ("[G]atekeeping simply refers to a regime of control over what content is allowed to emerge from the production processes in print and broadcast media; the controllers [journalists, editors, owners] of these media, in other words, control the gates through which content is released to their audiences.").

[158] *See generally id.* (describing the new phenomenon of gatewatching).

listeners."[159]

As scarcity shifts from speakers to listeners, the role of mass media is gradually demoted to mere gatewatching. To gatewatch is to not have control of the gate, but rather to rely more on the public as "selectors and filters of content."[160] In the digital media ecosystem, mass media is just one of many actors, albeit an important one, in shaping public discourse.[161]

The "bypass effect" brought about by social media has tremendous consequences for the way public debate holds the community together. Local elites—like the ones who fill the highest ranks in the state or control mass media organizations[162]—have a crucial role in enacting and enforcing speech norms in society.[163] The key distinction is not that pre-digital gatekeepers were *elites* enforcing their own norms, but simply that they were *local* community members steeped in shared norms. In general, people will not publish content they find unacceptable or that could invite social sanction. Those running local media were locals socialized within community standards. In aggregate, this meant gatekeeping organically reinforced prevailing social norms. Social media's detached global position disrupts this localized norm enforcement. Platforms are not inherently interested in reproducing particular community standards (potentially outside of U.S. norms), introducing a challenge. It is not a matter of pre-digital gatekeepers actively promoting their elite norms, but rather localized media personnel naturally tending to uphold norms

---

[159] Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547, 548 (2018).

[160] Enjolras & Steen-Johnsen, *supra* note 72, at 105.

[161] *See* Abiri, *supra* note 5, at 796 ("Mass media becomes merely one voice [albeit an extremely important one] in what reaches the public.").

[162] *See generally* EDWARD S. HERMAN & NOAM CHOMSKY, MANUFACTURING CONSENT: THE POLITICAL ECONOMY OF THE MASS MEDIA (2010) (discussing how local elites, such as high-ranking state officials or controllers of mass media, manipulate news to manufacture public consent. The authors' "propaganda model" illustrates how these power holders use media to perpetuate their interests, shaping public perception and influencing societal discourse, often against public interest).

[163] *See* Post, *supra* note 63, at 1042 ("In effect, newspapers create the conditions in which readers conceive themselves as a public, and hence as implicated in the formation of the public opinion that is essential to a democracy.").

they were socialized in themselves. Social media's global character severs this inherent local reinforcement.

All social-media platforms have some form of internal regulation (sometimes called "community standards"[164]) that they use to regulate communication among their users. These standards, however, differ from the kind of regulation that happens in the civil-society institutions we have encountered so far. These community standards are drafted in platform headquarters in California or Beijing or Singapore. These content-moderation rules are enforced by algorithms and thousands of human content moderators,[165] usually located in the developing world.[166] Despite efforts to make their enforcement more attuned to local communities and reach out to local stakeholders, social-media platforms' very attempt to enforce uniform speech norms on their global user base turns that mission into an oxymoron.[167]

Take our examples from Part II: insults, hate speech, and

---

[164] *See e.g.*, *Facebook Community Standards - Transparency Center*, FACEBOOK, https://transparency.fb.com/policies/community-standards [https://perma.cc/EWM2-HFAQ]; *Content Policy*, REDDIT, https://www.redditinc.com/policies/content-policy [https://perma.cc/J3BW-SE5M]; *Community Guidelines*, TIKTOK, https://www.tiktok.com/community-guidelines/en [https://perma.cc/2PEG-NSZ7].

[165] *See* Casey Newton, *The Trauma Floor: The Secret Lives of Facebook Moderators in America*, THE VERGE (Feb. 25, 2019), https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona [https://perma.cc/EZB7-WTFV].

[166] *See, e.g.*, Sana Ahmad & Martin Krzywdzinski, *Moderating in Obscurity: How Indian Content Moderators Work in Global Content Moderation Value Chains*, *in* DIGITAL WORK IN THE PLANETARY MARKET 77 (Mark Graham & Fabian Ferrari eds. 2022).

[167] Farhana Shahid & Aditya Vashistha, *Decolonizing Content Moderation: Does Uniform Global Community Standard Resemble Utopian Equality or Western Power Hegemony?*, PROC. 2023 CHI CONF. ON HUM. FACTORS COMPUTING SYS., Apr. 2023, at 1, 1 ("[T]he monolithic moderation systems often fail to account for large sociocultural differences between users in the Global South and users in the West.").

misinformation. We have explained how, in these three areas of communication, speech norms enforced by civil society institutions (and mass media in particular) help maintain the norms that hold a community together. Now let us look at how social media erodes the common understandings that make this possible.

Insults. Content moderation by platforms is not prepared to account for the incredible nuance that words have in different societies. The same word can be a terrible insult in one place and a friendly jab a few kilometers away—or even in the same place. In Argentina, the word *boludo* can be a very friendly vocative that indicates trust and horizontality, or an insult meant to undermine its recipient's intellectual ability or character. Distinguishing its different usages requires a deep immersion in the social life of Argentina; it is completely impossible for platforms to pen a rule that captures the distinction between when *boludo* is used as a friendly word and when it is a slur.[168]

Hate speech. Similarly, distinguishing between hate speech (with its underlying motivation of alienating some groups from the political community[169]) and strong, legitimate, disagreement, or unpalatable but genuine humor, is not easy. In many languages, for instance, some words have evolved out of their discriminatory origins. In Colombia, the word *marica* was originally a derogatory term directed at gay men, although currently is used alternatively as a generic vocative, such as "friend" or "dude," or as a generic insult, such as "dumb."[170]

---

[168] *See* Silvia Ramírez Gelbes & Andrea Estrada, *Vocativos "Insultivos" vs. Vocativos "Insultativos": Acerca del Caso de "Boludo"*, ANUARIO DE ESTUDIOS FILOLÓGICOS 335, 335 (2003) (noting that, while these usages are difficult to figure out in the abstract, both versions "are never misunderstood by the speakers").

[169] *See generally* JEREMY WALDRON, THE HARM IN HATE SPEECH (2012) (arguing that exclusion from the public is a central attribute of hate speech).

[170] Colombia Protest Case No. 2021-010-FB-UA, OVERSIGHT BOARD, https://www.oversightboard.com/decision/FB-E5M6QZGA [https://perma.cc/HY3W-KGSB] ("The Board sought expert input and public comments that confirmed that the word 'm**ica' has multiple meanings and can be used without discriminatory intent. However, there is agreement that its origins are homophobic, principally against gay men, even though its use has evolved to reportedly common usage in Colombia

After a series of street protests in 2021, Facebook removed a video of demonstrators who urged President Iván Duque to "stop playing *marica* on TV."[171] Was that a legitimate use of the word *marica*? Despite acknowledging that the word "has multiple meanings and can be used without discriminatory intent," Meta's Oversight Board seemed to think it was not a legitimate use, finding the word to maintain traces of its homophobic genesis.[172] However, the question in Colombia was far from settled.[173]

Misinformation. The identification of misinformation also depends on community norms, although in less obvious ways than the previous examples. Sometimes, whether an utterance constitutes misinformation depends wholly on whether it conforms to uncontestable facts or scientific truth, making moderation relatively straightforward. However, even a healthy public sphere is filled with unproven facts and theories,

---

to refer to a person as 'friend' or 'dude,' and as an insult equivalent to 'stupid,' 'dumb' or 'idiot.'").

[171] *Id.*

[172] *Id.* ("The Board notes that this evolution or normalization does not necessarily mean the term's usage is less harmful for gay men, as this casual use may continue to marginalize lesbian, gay, bisexual and transgender (LGBT) people and communities by implicitly associating them with negative characteristics."). It is worth mentioning that the Board ultimately decided to reverse Facebook's original decision to remove the video, although it did so because of the video's "newsworthiness allowance," under which "exceptional and limited harmful content would be permitted." *Id.*

[173] For a discussion in Colombia over the use of "marica," see, for example, Redacción BLU Radio, *¿Es discriminatorio usar la palabra "marica"? Debate en Mañanas BLU*, BLU RADIO (Blu Radio), https://www.bluradio.com/sociedad/es-discriminatorio-usar-la-palabra-marica-debate-en-mananas-blu. Some of the comments received by the Board also pointed out that the word "marica" does not automatically translate to a slur. *See, e.g.*, Fundación Karisma's Carolina Botero Comment no. PC-10194, in Colombia Protest Case No. 2021-010-FB-UA, OVERSIGHT BOARD, https://oversightboard.com/attachment/533957857868924 [https://perma.cc/DNB8-UHTQ].

fringe beliefs, exaggerations, and falsity.[174] To what extent these are acceptable depends heavily on community norms regarding the separation of truth and falsity.

Consider the burden of proof in everyday communication. The amount of information required to legitimately say something varies greatly according to the subject matter and the type of social values attached to it. For instance, how much evidence do I need to accuse someone publicly of a pernicious act? Does that amount change depending on the gravity of the act—whether it is a felony, or the fact that it was conducted without other witnesses than the victim? The response to these questions changes according to the social context, and deeply depends on social values and community norms.[175] In this way, social norms shape the epistemic conditions for what can and cannot be legitimately said in public. Through their professional norms, mass-media gatekeepers are accustomed to operationalizing these social judgments.

Take the informational chaos regarding the origins of COVID-19. In 2020, many social-media platforms flagged posts linking the SARS-CoV-2 virus to a lab leakage in Wuhan as misinformation.[176] Later, when President Biden's administration indicated that the lab-leakage theory was potentially credible, platforms ceased to apply that label to it.[177] Conservatives accused platforms of being hypocritical and

---

[174] *See* United States v. Alvarez, 567 U.S. 709 (2012) (generally stating that falsehood is not enough to deprive expression from First Amendment protection); *see also* Joanna M. Burkhardt, *History of Fake News*, 53 LIB. TECH. REP. 5, 5 (2017) ("Rumor and false stories have probably been around as long as humans have lived in groups where power matters.").

[175] This is even true when it comes to burden of proof standards in more formal settings. *See generally* Mirjan Damaska, *Rational and Irrational Proof Revisited*, 5 CARDOZO J. INT'L & COMP. L. 25, 34-36 (1997) (arguing that seemingly irrational proof practices in different cultures make more sense when viewed within their specific cultural contexts).

[176] F.D. Flam, *Facebook, YouTube Erred in Censoring Covid-19 'Misinformation'*, BLOOMBERG (June 7, 2021, 9:00 AM ET), https://www.bloomberg.com/opinion/articles/2021-06-07/facebook-youtube-erred-in-censoring-covid-19-misinformation [https://perma.cc/T9PY-97N3].

[177] Demetri Sevastopulo, *How Biden Came Round to the Wuhan Lab-Leak Theory*, FIN. TIMES (May 30, 2021), https://www.ft.com/content/923e0256-7f7e-43ef-b4fe-64c066b8b70b [https://perma.cc/8KLB-DLYS].

subservient to progressive thinking.[178] That, however, is an over-simplification. The reasoning behind platforms' decision to label lab-leakage theories as "misinformation" was not merely about facts, but also about values. Platforms would have hardly considered "misinformation" someone's statement that SARS-CoV-2 came down from a giraffe rather than a bat or a pangolin. The reason behind platforms' decision was probably that, in the context of the Trump presidency and in the midst of a catastrophe, the burden of proof needed to state a theory that could potentially instill Sinophobic sentiment in the American population was very high,[179] and speculations about a yet-unproven leakage from a Chinese lab had the potential of doing more harm than good.[180] This assessment

---

[178] Tara McKelvey, *Wuhan Lab-Leak Theory Fuels Trump Comeback Rally*, BBC NEWS (June 28, 2021), https://www.bbc.com/news/world-us-canada-57616323 [https://perma.cc/Y7YV-6GEV].

[179] Anti-Asian sentiments were reported to have spiked after the COVID-19 outbreak in the United States. *See* Neil G. Ruiz, Juliana Menasce Horowitz & Christine Tamir, *Many Black and Asian Americans Say They Have Experienced Discrimination Amid the COVID-19 Outbreak*, PEW RSCH. (July 1, 2020), https://www.pewresearch.org/social-trends/2020/07/01/many-black-and-asian-americans-say-they-have-experienced-discrimination-amid-the-covid-19-outbreak [https://perma.cc/U3WC-F5HA] (reporting that "[a] majority of Asian adults (58%) say it is more common for people to express racist or racially insensitive views about people who are Asian than it was before the coronavirus outbreak"). The same was true in the United Kingdom. *See* Jeffrey Howard, *Coronavirus Misinformation, Social Media and Freedom of Speech*, *in* POLITICAL PHILOSOPHY IN A PANDEMIC: ROUTES TO A MORE JUST FUTURE 167, 168 (Fay Niker & Aveek Bhattacharya eds., 2021) ("Misinformation about coronavirus is problematic for a number of reasons, but the most important reason is that it can lead to harm. . . . [T]he mistaken assumption that the Chinese are to be blamed for the virus seems at least partly responsible for the considerable spike in anti-Asian hate crime . . . ").

[180] Maxime Lepoutre, *Democratic Group Cognition*, 48 PHIL. & PUB. AFF. 24, 64 (2020) (reconstructing Heather Douglas' argument that "when determining whether their evidence for d is sufficient to warrant asserting d, responsible scientists should weigh the costs of false positives [asserting

might have been wrong, but there is nothing outrageous about it. Rather, it is the customary activity of civil-society gatekeepers in public discourse. The problem is that, platforms being global, this assessment was done taking into consideration U.S. values and context. A gatekeeper in Cameroon, China, or Mexico would have weighed different considerations, and may have reached a different decision.

As these examples show, global platforms have eroded the power of community elites to gatekeep the public sphere according to relevant community norms. This subtraction of power from the elites did not accrue to other members of the community, but rather to global elites disconnected from the various community norms of the societies in which platforms operate. These global elites enact and implement their own norms for what they call "content moderation." In doing so, they allow speakers to bypass gatekeepers, preventing them from exerting their traditional function in the local public sphere. We call this phenomenon the "bypass effect."

V. Localism Versus Universalism

The foundational political challenge posed by the bypass effect forces states to respond, accepting some of the features of the new communicative ecosystem and rejecting others. We can place these responses on a continuum: On the one end, *universalism* embraces the bypass effect, and finds virtue in the global nature of the digital public sphere promised by platforms. It strives for a global square in which citizens of the world come together to deliberate on the same terms, governed by global speech norms. On the other end, *localism* fiercely resists the bypass effect, opposing the overtake of the local public square by alien forces such as globalized, U.S.-based platforms. Most real-world regulations fall somewhere between these two poles, embracing or accepting some speech regulation by global actors while maintaining or conceding regulatory power by local actors. Before delving into the complexities of the real world, however, let us take a closer look at the two ends of the spectrum.

---

d when it is false] against the costs of false negatives [not asserting d when it is true]. This weighing, in turn, necessarily relies on normative judgments about the value or disvalue of different outcomes").

A. Platform Universalism

Universalists dispute the notion of the bypass effect posing a problem. They embrace the Internet's transnational nature as a "global public good" that transcends the bounds of state sovereignty.[181] The digitalization of media, from their perspective, is at the forefront of the globalization process, bearing immense prospects for scientific collaboration, cultural interchange, and the fostering of a more interconnected global community. Eventually, the universalist will tell us, the Internet will create a global public sphere. But in the meantime, the bypass effect is liberating rather than a problem. We are better off without these cumbersome, oppressive gatekeepers that only prevented truth and freedom from prevailing.

In its purest version, therefore, universalism is deeply motivated by a cosmopolitan ideology, one which posits that humans are citizens of the world and that national allegiances are relics of the past.[182] This type of ideology was common among early Internet activists, who put it in the most epic terms: "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us

---

[181] Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1206 (2017) ("Currently the Internet is mostly governed by the values of the least censorious regime—that of the United States. If nation states can enforce global filtering, blocking, and delinking, the Internet will eventually be governed by the most censorious regime. This will undermine the global public good of a free internet.").

[182] *See generally* Pauline Kleingeld & Eric Brown, *Cosmopolitanism*, STAN. ENCYC. PHIL. (Oct. 17, 2019), https://plato.stanford.edu/entries/cosmopolitanism [https://perma.cc/MQS3-WFVP] (defining "cosmopolitanism" as a set of ideologies sharing the belief that "all human beings, regardless of their political affiliation, are (or can and should be) citizens in a single community," and pointing out that some strains of cosmopolitanism "den[y] the existence of special obligations to local forms of political organizations").

alone. You are not welcome among us. You have no sovereignty where we gather."[183]

Universalism is part of the official ideology spread by social media companies. Tech billionaire after tech billionaire claim to be attempting to unite the world community,[184] create a global consciousness,[185] and establish a global public square.[186] This new global public sphere is to be governed by common rules, which are usually labeled "community standards."[187] The "community" here, apparently, is the whole of humanity.[188]

To produce such a community, they, in principle, should apply their "community standards" to each user no matter where they live or access the web. There is an inherent tension in such a project: On the one hand, these platforms are committed to giving people equal access to expressing themselves, whether rich or poor, from the Global South or North.[189] On the other hand, when applying a single set of norms to everyone, one must ask whose norms are being applied, and to whose community.[190]

In a different world, universal speech norms are adopted after a process of global deliberation that takes into

---

[183] Barlow, *supra* note 1.

[184] Mark Zuckerberg, *Building Global Community*, FACEBOOK (May 5, 2021), https://www.facebook.com/notes/3707971095882612 [https://perma.cc/L7QY-R3U4] ("Progress now requires humanity coming together not just as cities or nations, but also as a global community.").

[185] *See* Jack Dorsey's definition *supra* note 3.

[186] Jillian C. Jork, *Elon Musk Doesn't Know What It Takes to Make a Digital Town Square*, MIT TECH. REV., https://www.technologyreview.com/2022/10/29/1062417/elon-musk-twitter-takeover-global-democracy-activists [https://perma.cc/E7K7-GGAE] ("Musk says he acquired Twitter because it is "important to the future of civilization to have a common digital town square, where a wide range of beliefs can be debated in a healthy manner without resorting to violence."").

[187] *Facebook Community Standards*, *supra* note 164.

[188] On the ambiguous use of the term "community" by social media platforms, see generally Emiliano Vitaliani, Morena Schatzky & Matias Gonzalez, *"Community" in the Digital Realm*, YALE ISP KNIGHT SERIES, Sept. 12, 2023, at 1.

[189] *See Facebook Community Standards*, *supra* note 164 ("Our Community Standards apply to everyone, all around the world, and to all types of content.").

[190] *See generally* Shahid & Vashistha, *supra* note 167 (providing an account of how this universalism feels in the Global South).

consideration the interests and opinions of everyone who takes part in global communication.[191] This is not the world we live in. When applied in the context of real-world politics and technology, universalism tends to transform into *imperialism*. The norms adopted as universal are *someone*'s norms, and that someone is usually an employee of a California-based tech giant, influenced both by elite U.S. culture and U.S. free-speech regulation.[192] When it comes to cultural influence worldwide, California Big Tech comes to finish what Hollywood started.[193]

In the real world, there are two political realities that push countries into the universalist direction: imperialism and impotence. It is easy for the United States to be universalist because of the values and identity of platforms align with their own:[194] Section 230, an iconic piece of universalist legislation, would probably look very different if most hegemonic platforms had been Chinese from the start.[195] However, most countries that adopt universalist stances do so because of sheer impotence to meaningfully regulate platforms. For these small,

---

[191] *See generally* Nahuel Maisley, *The International Right of Rights? Article 25 (a) of the ICCPR as a Human Right to Take Part in International Law-Making*, 28 EUR. J. INT'L L. 89 (2017) (discussing the right to participate in the creation of global norms).

[192] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1616-22 (2017) (arguing that Facebook, Twitter, and YouTube's free speech baseline is grounded in First Amendment values).

[193] *See generally* CHRISTIAN FUCHS, CRITICAL THEORY OF COMMUNICATION: NEW READINGS OF LUKÁCS, ADORNO, MARCUSE, HONNETH AND HABERMAS IN THE AGE OF THE INTERNET (2016) (drawing attention to the growth and dominance of American-based platforms as signaling a new phase of media, informational, and cultural imperialism).

[194] Some go so far as to name this alignment "Platform Imperialism." *See* Dal Yong Jin, *The Construction of Platform Imperialism in the Globalization Era*, 11 TRIPLEC 145, 145 (2013).

[195] 47 U.S.C. § 230 (known as "Section 230") is universalist because it is so laissez faire towards the platforms. For a good account of its hand-off approach see Tarleton Gillespie, *Platforms Are Not Intermediaries*, 2 GEO. L. TECH. REV. 198, 204 (2018).

poor, or not very tech-savvy states, universalism might not strike a great balance, but what option do they have?

Universalism, therefore, looks very different from Palo Alto than it looks from El Salvador or Myanmar. For those who carry U.S. interests at heart, it may make sense to let companies do as they please, since they are going to naturally pursue and enforce U.S. values and speech norms. Undeniably, there are instances when U.S. legislators demonstrate their readiness to expose social-media corporations to public scrutiny, particularly when these entities meddle with matters of governmental significance.[196] However, the intensity of this pressure pales in comparison to the force exerted upon platforms originating from a foreign country.[197] We turn to this type of *localist* pressure next.

B. Platform Localism

From the localist perspective, the platform-driven Internet is a destabilizing element that requires domestic regulation. The transnational character of the Internet is not seen as an asset, but as a complication needing rectification, if not reversal. This does not suggest that localists completely dismiss the advantages of an open Internet, particularly the economic windfall of global communication. Yet, the concept of a "global public square" bears little relevance or appeal to them.

Localists are not necessarily driven by racial bias or xenophobia. Their perspective is rooted in the appreciation for local culture and political infrastructures and an inherent desire to shield these from the broad strokes of a globalized internet platform. As such, internet localism can be interpreted as an element of a more general resistance to cultural globalization.[198]

---

[196] *See, e.g.*, *Zuckerberg and Dorsey Face Harsh Questioning from Lawmakers*, N.Y. TIMES (Nov. 17, 2020), https://www.nytimes.com/live/2020/11/17/technology/twitter-facebook-hearings [https://perma.cc/3R2X-QP2H].

[197] *See, e.g.*, Alex Cranz, *That TikTok Hearing Was Pretty Messed Up, Right?*, THE VERGE (Mar. 24, 2023, 4:58 PM ET), https://www.theverge.com/2023/3/24/23654831/tiktok-congressional-hearing-xenophobia-china [https://perma.cc/8D65-5HY9]. *See also infra* notes 212-218 and accompanying text.

[198] For a classical popular account of this critique, see BENJAMIN R. BARBER, JIHAD VS. MCWORLD (2010).

Universalists might appear nonchalant towards the bypass effect, but from the localists' perspective, it is a catastrophe in slow motion. The weakening of local gatekeepers—those who historically have shaped public discourse while upholding community norms—poses a threat to societal stability and cohesion, or at the very least, to the established mechanisms that communities have employed to preserve their equilibrium. This results in an almost instinctive rejection of transnational social media.

The prospect of a community losing control over its communicational environment can legitimately preoccupy any regime, be it an authoritarian or a democratic one. But, practically, localism is more easily implemented by illiberal states. In liberal communities, which have largely delegated control over the public sphere to civil-society institutions,[199] fighting back against the bypass effect becomes extremely difficult. By nature, civil-society institutions do not have central coordination, so it is not possible for them to agree on a common action plan.[200] Furthermore, even if they did, they lack the coercion tools necessary to prevent transnational actors from reaching citizens directly, at least if citizens are sufficiently attracted to them. A local newspaper's soft power will have a hard time against Facebook or TikTok.

Localism, therefore, quickly turns into *statism*. Effective localist reactions to platform power are found primarily in communities where the state is the main actor governing the public sphere. They have both the motive and the opportunity. The bypass effect is more of a threat to authoritarian regimes that are invested in controlling the public sphere not only to keep local cohesion, but also to maintain their own rule. To see how platforms can be a threat to authoritarian regimes, just

---

[199] *See supra* Section III.B.

[200] MICHAEL EDWARDS, CIVIL SOCIETY 43 (2004) ("The reality is that norms vary between different associations in the same society or culture and between different cultures and societies – not exactly rocket science, but crucial to a clear view of potential pathways to social progress.").

think of the Arab Spring.[201] And indeed, particularly when we turn our attention to proficient authoritarian states such as China, they have demonstrated the capacity to effectively suppress platforms.[202] This ability stems from their control over the machinery of state coercion—potentially the singular force potent enough to compel platforms into retreat.

It makes sense, then, that the most emblematic and effective localist internet regime is the Chinese one. The Chinese Internet is, to a great degree, already split from the globalized Internet.[203] Their approach takes two forms: on the level of platform-based Internet (web 2.0), there is a complete split: major global corporations were either invited to leave the Chinese market (Google,[204] LinkedIn[205]) or outright banned (Facebook,[206] Twitter[207]). At the same time, most Chinese platforms make little to no effort to appeal to foreign users (some do not even have an English version!), and many require

---

[201] *See generally* ZEYNEP TUFEKCI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 13, 27 (2017) ("For decades, authoritarian states in Egypt, Tunisia, and other countries in the Arab world built up extensive control and censorship of the mass media. . . . Digital technologies, along with the satellite TV channel Al Jazeera, changed this situation. . . . Thanks to a Facebook page, perhaps for the first time in history, an internet user could click yes on an electronic invitation to a revolution. Hundreds of thousands did so . . . .").

[202] LinkedIn was the last U.S.-based social media platform to get pushed out of China. *See* Karen Weise & Paul Mozur, *LinkedIn to Shut Down Service in China, Citing 'Challenging' Environment*, N.Y. TIMES (Oct. 14, 2021), https://www.nytimes.com/2021/10/14/technology/linkedin-china-microsoft.html [https://perma.cc/3779-B7U7].

[203] *See, e.g.*, Daniel Anderson, *Splinternet Behind the Great Firewall of China: Once China Opened Its Door to the World, It Could Not Close It Again.*, 10 QUEUE 40 (2012).

[204] Matt Sheehan, *How Google Took on China—and Lost*, MIT TECH. REV. (Dec. 19, 2018), https://www.technologyreview.com/2018/12/19/138307/how-google-took-on-china-and-lost [https://perma.cc/Y23W-CMYF].

[205] Weise & Mozur, *supra* note 202.

[206] Alyssa Abkowitz et al., *Facebook Is Trying Everything to Re-Enter China—and It's Not Working*, WALL ST. J. (Jan. 30, 2017, 10:45 AM ET), https://www.wsj.com/articles/mark-zuckerbergs-beijing-blues-1485791106 [https://perma.cc/Y7ND-ADZ4].

[207] David Bamman et al., *Censorship and Deletion Practices in Chinese Social Media*, 17 FIRST MONDAY, Mar. 4, 2012, at 1 ("Twitter and Facebook were blocked in China in July 2009 after riots in the western province of Xinjiang.").

a Chinese phone number for registration.[208] The fact that TikTok is separated from its Chinese identical twin (Douyin) is emblematic of this split in Web 2.0.[209]

When moving down to the protocol-based Internet (web 1.0) the Chinese Internet is not hermetically sealed—Chinese citizens can visit foreign websites and exchange emails. However, their experience is constantly moderated by the so-called Great Firewall: a massive filtering system that prevents users inside of China from viewing sites deemed harmful to the nation.[210] This dual model allows the Chinese state apparatus to enjoy some of the economic and scientific benefits of an open Internet while at the same time exerting an incredible amount of control over internal and external communication. As all of this suggests, there are various ways in which a competent illiberal state can tame the bypass effect.

Although localism, as an institutional reality, can likely only take full form in a competent illiberal regime, the basic motivations of political stability and prevention of foreign interference are widespread. Sometimes, it merely takes a hypothetical threat to national security to convert the most dedicated universalist into a fervent localist.[211] Many U.S. lawmakers, for instance, became suddenly worried about

---

[208] *See e.g.*, Jennifer Pak, *How Easy Is it for Foreign Visitors to Use Chinese Apps?*, MARKETPLACE (Nov. 21, 2022, 5:30 AM), https://www.marketplace.org/2023/11/21/how-easy-is-it-for-foreign-visitors-to-use-chinese-apps [https://perma.cc/ZY5N-G5VC].

[209] Jessie Yeung & Selina Wang, *TikTok is Owned by a Chinese Company. So Why Doesn't It Exist There?*, CNN BUS. (Mar. 24, 2023, 7:58 AM EDT), https://edition.cnn.com/2023/03/24/tech/tiktok-douyin-bytedance-china-intl-hnk/index.html [https://perma.cc/64U3-ZQL].

[210] For an exploration of China's Internet filter, nicknamed the "Great Firewall," see generally Roya Ensafi et al., *Analyzing the Great Firewall of China over Space and Time.*, 2015 PROC. PRIV. ENHANCING TECHNOL. 61 (2015); MARGARET ROBERTS, CENSORED: DISTRACTION AND DIVERSION INSIDE CHINA'S GREAT FIREWALL (2018).

[211] For a description of how existential threats can spur a "rally 'round the flag," see Gilad Abiri & Sebastián Guidi, *The Pandemic Constitution*, 60 COLUM. J. TRANS'L L. 68, 116-18 (2012).

whose values social-media platforms uphold only when a Chinese-owned platform became prominent.[212]

In the 2023 Congressional hearings with TikTok's CEO Shou Chew, the formal excuse for the inquiry was potential undue Chinese Communist Party access to Americans' data. However, a lot of the language used across party lines suggests that behind this justification lies a deep concern about foreign interference with U.S. public debate. In the Chair's opening remarks, the problem was not merely that TikTok had access to data, but that the Chinese government would use the data "as a tool to manipulate America as a whole."[213] Fundamentally, as the Chair put it, the problem was that "[w]e do not trust TikTok will ever embrace American values, values for freedom, human rights, and innovation"[214] and therefore it "should be banned."[215] Texas Senator Ted Cruz followed a similar line of argument: "TikTok is a Trojan horse the Chinese Communist Party can use to influence what Americans see, hear, and ultimately think."[216]

This concern was not limited to the Trumpian wing of the Republican Party, with its distinctively isolationist instincts. For instance, Scott Galloway, one of the leading media voices

---

[212] Actually, American-owned platforms became subject to Congressional scrutiny before that, but a good part of the objections they received were related, precisely, to the accusation of helping and abetting foreign powers to interfere in American public debate–that is, they were accusing platforms of not being American enough. *See, e.g.*, Miles Parks & Philip Ewing, *Foreign Interference Persists and Techniques Are Evolving*, *Big Tech Tells Hill*, NPR (June 18, 2020, 3:09 PM), https://www.npr.org/2020/06/18/880349422/foreign-interference-persists-and-techniques-are-evolving-big-tech-tells-hill [https://perma.cc/H33Z-N87W]; T. S. Allen Rodriguez Stephen, *To Protect Democracy, Protect the Internet*, FOREIGN POL'Y (July 14, 2020, 4:05 PM), https://foreignpolicy.com/2020/07/14/united-states-election-interference-illegal-social-media [https://perma.cc/2DUW-T6GT].

[213] *Transcript: TikTok CEO Testifies to Congress*, TECH POL'Y PRESS, https://techpolicy.press/transcript-tiktok-ceo-testifies-to-congress [https://perma.cc/5KZ7-KP6S].

[214] *Id.*

[215] *Id.*

[216] *See* Drew Harwell & Elizabeth Dwoskin, *As Washington Wavers on TikTok, Beijing Exerts Control*, WASH. POST (Oct. 30, 2022, 1:33 PM), https://www.washingtonpost.com/technology/interactive/2022/bytedance-tiktok-privacy-china [https://perma.cc/PJG6-VMJQ].

on technology and an overall centrist, wrote that "[t]he tip of China's propaganda spear is TikTok, which has a direct connection to the midbrain of a billion people, including nearly every U.S. teenager and half their parents. . . . [N]ow China commands the most powerful propaganda tool."[217] Democratic Senator Michael Bennet was also of a similar opinion "[China is] now hypnotizing us with TikTok."[218] As *The Verge* wrote in a subtitle after the hearings, "[i]nstead of asking actual important questions related to how TikTok does business and uses the data it gathers from users, Congress was focused on being xenophobic."[219] Localist instincts and policies, we see, are not limited to autocracies.

VI. The Middle Path? Two Forms of Federalism

Universalism and localism each fail where the other succeeds. Universalism is relatively easy to implement: if unleashed, market forces are likely to deliver universal content-moderation policies. However, this inaction can have destabilizing consequences for our global order, which is still reliant on local nation-states. Localism, on the other hand, has the potential to counter the bypass effect successfully, but at the price of risking authoritarianism and losing the tremendous scientific, cultural, and economic benefits of the open Internet.

Most countries, it seems, want to stay on the Internet and leave at the same time. They seek to reap the advantages of the centralized scale economy of the platform Internet while reserving certain decisions for their political communities.

---

[217] Scott Galloway, *Why TikTok is a Trojan Horse for Espionage and Propaganda,* MEDIUM (July 11, 2022), https://medium.com/@profgalloway/tiktok-trojan-stallion-64169f194d87 [https://perma.cc/DWF6-48MD].

[218] Press Release, Michael Bennet, U.S. Senator, Bennet Calls for Ban on TikTok: "The Sooner We Ban This, The Better" (Mar. 23, 2023), https://www.bennet.senate.gov/public/index.cfm/2023/3/bennet-calls-for-ban-on-tiktok-the-sooner-we-ban-this-the-better [https://perma.cc/R8TX-ZPBY].

[219] Cranz, *supra* note 197.

They want to have the final decision on some issues, but understand they should concede on others. Ultimately, they want a say without full secession.

All these features make these countries' attitudes towards platform power akin to traditional federal constitutional arrangements, in which power and authority are divided between a central government and smaller, constituent political units. In traditional federal arrangements, this division of power seeks to balance the benefits of a unified, centralized government with the preservation of regional autonomy, allowing for a potentially more effective distribution of power and resources while also respecting local interests and values.[220]

Of course, we should keep the analogy at bay. We do not mean to say that every feature of constitutional federalism can apply to global platform governance.[221] But we mean more than a creative metaphor:[222] constitutional and platform federalism are motivated by the same contradictory desire to leave and stay at the same time;[223] to "unite for common purposes yet

---

[220] *See* DANIEL J. ELAZAR, EXPLORING FEDERALISM 5-6 (1987) ("As a political principle, federalism has to do with the constitutional diffusion of power so that the constituting elements in a federal arrangement share in the processes of common policy making and administration by right, while the activities of the common government are conducted in such a way as to maintain their respective integrities. Federal systems do this by constitutionally distributing power among general and constituent governing bodies in a manner designed to protect the existence and authority of all.").

[221] *See e.g.*, Thomas E. Kadri, *Juridical Discourse for Platforms*, 136 HARV. L. REV. F. 163, 198 (2022) ("To be clear, federalism analogies go only so far because of salient differences between public and platform governance.").

[222] *See* ELAZAR, *supra* note 220 at 15-16 ("[F]ederalism is . . . a value concept—a term that carries with it an essence, which is interpreted in a variety of ways under different circumstances as long as they adhere to the essentials of the concept so that they serve to allow people to "hone in" on a particular set of deep structural meanings. As a value concept, it does not have a once-and-for-all-time precise definition in the usual scientific sense, although it can be and is defined operationally in well-accepted ways.").

[223] *See id*. at 11 ("In most if not all of these cases, whether they know it or not, the various parties have arrived at the point which the late Martin Diamond described as the classic position of federalism—the position expressed by the song that Jimmy Durante, the American comedian, belted out in the film, The Man Who Came to Dinner: 'Did you ever have the feeling that you want to go, and the feeling that you want to stay?' That is the classic problem for which federalism, as a technology, was invented.").

remain separate to preserve their respective integrities."[224] In both cases, federalism attempts to "supplement[] but does not seek to replace or diminish prior organic ties where they exist."[225]

The federalist view of the Internet is ambivalent and value-pluralist. On the one hand, it sees great value in the existence of transnational communication, for cultural and economic reasons. On the other hand, it believes that nation-states (or supranational entities such as the European Union) have a crucial role in battling the harms caused by the bypass effect. Federalists are skeptical that global platforms can be responsive to the values and interests of their citizens and see it as the state's role to force the platforms to take local needs into account. As a hybrid model, federalism is plagued with the question of balancing: if states regulate platforms too heavily, they shift into a localist model, but if they intervene too little, they are *de facto* allowing for a universalist model to emerge.

A federalist attitude towards platform regulation, however, can take at least two different forms, depending on the means states deploy to channel their wishes. Most countries have adopted what we call *state federalism*—a variation of federalism that directly dictates content-moderation policies to digital platforms. Threatened by the dilution of communitarian speech norms by the dissolving forces of globalization, they take up the task of restoring those norms themselves.

We propose an alternate view, which we call *civil society federalism*. In civil society federalism, the state also deploys various legal instruments, but, unlike in state federalism, it does so not to directly enforce community norms. Instead, it reinstitutes civil-society institutions as the gatekeepers of speech norms. In this way, civil society federalism seeks to restore the balance that existed before the bypass effect. Thus, such a regime can ensure some sensitivity to local idiosyncrasy and sentiment, while not exercising the power to control public

[224] *Id.* at 33.
[225] *Id.* at 12.

discourse directly in a way typical of illiberal regimes. We subsequently examine these two modes of platform federalism.

A. State Federalism

Most countries adopt what we call here state federalism; a variant of platform federalism in which governments directly dictate the speech norms that they want platforms to enforce within their borders. Unlike localism, state federalism permits an overwhelming majority of speech on platforms to stay uncensored. However, unlike universalism, it draws some lines that are not to be transgressed. Direct regulation of platforms' content-moderation policies allows state federalism to benefit from globalization without entirely losing communal identity.

State federalists detect the bypass effect and its pernicious implications for public debate and societal cohesion. They see information gatekeepers in their communities losing all ability to channel public debate in a trustworthy way. They notice the spread of fake news and hate speech and panic about the impotence of existing arrangements for preventing them. In this crossroad, they choose to replace civil-society actors with hardcore state regulation.

To a large extent, Internet state federalism is already a reality. In recent years, all the major digital platforms established a separate system on top of their universalist community standards to enforce local laws over content.[226] Some countries have also experimented with regulations that impose strict duties on platforms to regulate users' speech. Let us examine three core groups of such regulations. These groups share the same concerns about platform power, and they fail in different ways to address the issue.

1. Direct Law Enforcement

The quintessential state-federalist approach is to mandate

[226] *See* Chinmayi Arun, *Facebook's Faces*, 135 HARV. L. REV. F. 236, 239 (2021) ("Facebook mostly separates its engagement with states from its engagement with its users in the context of decisions about speech on its platform. It does this by using two different systems to regulate content. States expect the company to follow their laws and are interested in the company's implementation of those laws. The local-law system for determining unlawful content leads to the blocking of content in the state where it is illegal, while the content is available and visible on Facebook elsewhere.").

platforms to censor directly any speech that is already prohibited by national law. The most prominent example comes from Germany. In 2017, Germany enacted the Network Enforcement Act (NetzDG), which requires global platforms to remove within Germany content that violates the German Criminal Code, such as hate speech or Holocaust denial.[227] NetzDG is generally respected by platforms.[228]

While requiring the enforcement of an already existing criminal code on digital platforms may sound like maintaining the status quo, it actually increases state power over speech dramatically.[229] This becomes evident in authoritarian states,[230] but remains true even in liberal ones.[231] The German Criminal

---

[227] *See* Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerkdurchsetzungsgesetz-NetzDG] [Network Enforcement Act], Sept. 1, 2017, Bundesgesetzblatt, Teil I [BGBL I], at 3352 (Ger.), https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html [https://perma.cc/U8BK-5E9R].

[228] *See e.g.*, GOOGLE, TRANSPARENCY REPORT, https://transparencyreport.google.com/netzdg/youtube [https://perma.cc/V2SH-HJBB]. The NetzDG has been superseded by the E.U. Digital Services Act.

[229] Jack Balkin has distinguished between "old school" speech regulation, which proceeds mostly through civil damages and criminal fines, and "new school" speech regulation, which proceeds through regulating Internet infrastructure and is therefore more pervasive and relatively permissive of prior restraint through "collateral censorship." *See* Balkin *supra* note 181, at 1172-82.

[230] For instance, Venezuela has passed a "constitutional law" against hate, which prohibits various widely-defined forms of discrimination and "hatred" on social media. The law requires platforms to "take measures to prevent the diffusion of these messages" and "take down immediately any message [contrary to this law]." *See* Ley Constitucional contra el Odio, por la Convivencia Pacífica y la Tolerancia [Constitutional Law Against Hate, for Pacific Coexistence and Tolerance], CXLV-I Gaceta Oficial de la República Bolivariana de Venezuela No. 41.274 (Nov. 8, 2017), § 14, http://historico.tsj.gob.ve/gaceta/noviembre/8112017/8112017-5076.pdf [https://perma.cc/2U3K-TSN5].

[231] In 2020, the French Parliament enacted the so-called "Avia Law," regulating hate speech on social media platforms. The law required

Code, as any criminal code, is not meant to have the kind of automatic, overarching enforcement power wielded by platforms.[232] Criminal laws have an expressive function that is fulfilled even in the absence of enforcement,[233] and it is hard to know how stringently legislatures expect them to be enforced. Frictions in enforcement constantly allow "public opinion [to] rewrite[] the terms of the criminal code."[234]

This is particularly salient in speech, an area in which "subsequent imposition of liability" is preferred to prior restraint.[235] The "breathing space" created by less-than-perfect enforcement allows for a dialectic relationship between social

---

platforms to take down widely-defined illegal content within twenty-four hours after any user flags it (one hour if the content is related to child sexual abuse or terrorism). The law was heavily criticized by human rights organizations. *See French Avia Law Declared Unconstitutional: What Does This Teach Us at EU Level?*, EDRI (June 24, 2020), https://edri.org/our-work/french-avia-law-declared-unconstitutional-what-does-this-teach-us-at-eu-level [https://perma.cc/9EY3-JXY2]. It was ultimately struck down by the Constitutional Council who decided that the law "undermines freedom of expression and communication in a way that is not necessary, adapted nor proportionate," since it gave platforms an incentive to remove flagged content regardless of its illegality in order to avoid liability. *See* Conseil constitutionnel [CC] [Constitutional Council] decision No. 2020-801 DC, June 18, 2020, https://www.conseil-constitutionnel.fr/decision/2020/2020801DC.htm [https://perma.cc/FNQ9-W9A2].

[232] *See supra* note 126 and accompanying text.

[233] *See* Cass R. Sunstein, *On the Expressive Function of Law*, 124 PENN. L. REV. 2021, 2032 (1996) ("These cases arise when the relevant law announces or signals a change in social norms unaccompanied by much in the way of enforcement activity. Consider, for example, laws that forbid littering and laws that require people to clean up after their dogs. In many localities such laws are rarely enforced through the criminal law, but they have an important effect in signaling appropriate behavior and in inculcating the expectation of social opprobrium and, hence, shame in those who deviate from the announced norm. With or without enforcement activity, such laws can help reconstruct norms and the social meaning of action.").

[234] Marc L. Miller & Ronald F. Wright, *The Black Box*, 94 IOWA L. REV. 125, 148 (2008) ("Chief prosecutors, who must face the voters every few years, understand that some legally valid applications of the criminal law would be political suicide. Current public opinion constantly rewrites the terms of a criminal code drafted by legislatures over many decades.").

[235] Organization of American States, American Convention on Human Rights, Nov. 22, 1969, O.A.S.T.S. No. 36, 1144 U.N.T.S. 123, § 13.2.

customs and legal norms.[236] Only if some nominally illegal speech is allowed to exist, even if ultimately punished, can society have enough exposure to it to decide whether to permit it eventually. When the regulation of speech is partially delegated to civil-society institutions, this type of limit-testing can flourish.[237] The kind of direct regulation promoted by a law like NetzDG, combined with a foreseeable tendency by platforms to engage in over-enforcement to avoid liability,[238] impedes the social evolution of norms.

2. Fake News Prevention

Pre-digital gatekeepers enforced not only civility norms concerning appropriate speech, but also epistemic norms concerning the adequate level of trustworthiness a piece of information should have before gaining legitimacy in public discourse.[239] As the bypass effect prevents this type of epistemic

---

[236] *See supra* note 129 and accompanying text.

[237] This is very similar to the situation on Reddit. Imagine a new subreddit emerges with minimal moderation. Members begin posting hateful content, causing outrage within the broader Reddit community. Instead of completely shutting down the subreddit, Reddit administrators implement penalties for the most offensive posts. This gives the community and Reddit time to negotiate what constitutes unacceptable speech, fostering a dialogue that avoids overly restrictive pre-censorship.

[238] *See* Jack M. Balkin, *Old School/New School Speech Regulation,* 127 HARV. L. REV. 2296, 2308-14 (2014) (arguing that in new school speech regulation "[t]he state creates incentives for [platforms] to overcensor"). Human rights organizations have levied this critique against German NetzDG and the French Avia Law. *See supra* note 231 and accompanying text; *see also Germany: Flawed Social Media Law. NetzDG is Wrong Response to Online Abuse*, HUM. RTS. WATCH (Feb. 14, 2018, 12:01 AM), https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law [https://perma.cc/L7TR-45JN] ("[NetzDG] is vague, overbroad, and turns private companies into overzealous censors to avoid steep fines, leaving users with no judicial oversight or right to appeal.").

[239] *See* Abiri & Buchheim, *supra* note 76, at 65-66 ("[In the pre-digital world, media] elites, therefore, functioned as effective–though not always beneficial–gatekeepers for the publishing and spread of information. This gave them the power to enforce certain epistemic norms. They could

control by civil-society institutions when it comes to information hosted on platforms, it is only natural that some states have tried to take on this dangerous function of acting as sommeliers of truth and falsity.

Many governments are contemplating various laws and regulations against misinformation. France has passed an election misinformation regulation aimed at digital platforms, which empowers courts, the French Audiovisual Council, and Internet service providers to take different measures to flag or even remove fake information.[240] A similar law was passed in May 2019 in Singapore, a developed but authoritarian country. This law grants the Singaporean government authority to order corrections to, removals of, or access restrictions on online content deemed false and harmful to the public interest.[241] Lawmakers in two of the largest democracies in the world have proposed similar, albeit wider, regulations.[242] As these initiatives proliferate, many are voicing concerns about governments becoming the final arbiters of truth.[243] Fake news

---

maintain relatively common standards for verifying facts and widely shared trust in epistemic authorities.").

[240] *See* Loi organique n° 2018-1201 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [Organic Law no. 2018-1201 of December 22, 2018 on the Fight Against Manipulation of Information], Journal Officiel de la République Française n° 0297 (Dec. 23, 2018).

[241] *See* Protection from Online Falsehoods and Manipulation Act 2019 (Oct. 2, 2019) (Sing.). *See also* Kirnsten Hun, *Want to Criticize Singapore? Expect a 'Correction Notice'*, N.Y. TIMES (Jan. 21, 2020), https://www.nytimes.com/2020/01/21/opinion/fake-news-law-singapore.html [https://perma.cc/6L5K-ZJFV] ("Since the law came into force in October, the government has invoked it five times, and there is now reason to fear that the law is, instead, a tool to quiet dissent.").

[242] *See e.g.*, Projeto de Lei [Bill] no. 2630, de 2020 (Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet) [Brazilian Law for Freedom, Responsibility, and Transparency in the Internet], Senator Alessandro Vieira [Br.]; Menaka Doshi, *India Plans Sweeping Overhaul of Digital Law as Market Swells*, BLOOMBERG (May 24, 2023), https://www.bloomberg.com/news/articles/2023-05-24/india-plans-sweeping-overhaul-of-digital-law-as-market-swells [https://perma.cc/2A2U-SDLJ].

[243] *See e.g.*, Flemming Rose & Jacob Mchangama, *History Proves How Dangerous It Is to Have the Government Regulate Fake News*, WASH. POST (Oct. 3, 2017, 12:40 PM), https://www.washingtonpost.com/news/theworldpost/wp/2017/10/03/history

poses an obvious problem for governance and social cohesion.[244] However, state-enforced veracity tests are precisely the type of regulations that the liberal canon is terrified of, and for good reason.[245]

3. Limits on Content Moderation

State federalism does not always require platforms to engage in more content moderation; it can also demand that they do significantly less. Recently, both Texas and Florida passed legislation that virtually forbids social-media platforms from removing any content that is not illegal (which, under current First Amendment doctrine, is a very narrow category). This legislation is currently in the Supreme Court's docket.[246]

-proves-how-dangerous-it-is-to-have-the-government-regulate-fake-news [https://perma.cc/Y4NH-FXCS] ("[I]t is . . . crucial for a genuine liberal democracy that no one, and particularly no official institution, gets to determine the truth in matters of politics and public debate."); *Censorious Governments Are Abusing "Fake News" Laws*, ECONOMIST (Feb. 13, 2021), https://www.economist.com/international/2021/02/13/censorious-governments-are-abusing-fake-news-laws [https://perma.cc/8MBP-NMTN] (explaining and criticizing how the COVID-19 pandemic gave authoritarian "governments . . . an excuse to crack down on their critics using the pretext of restricting the spread of fake news.").

[244] *See* Abiri & Buchheim, *supra* note 76, at 62-63 ("[W]hat is new and challenging about the spread of digital falsities is that they have developed the potential to divide our societies into separate epistemic communities. . . . While a specific instance of untrue belief can be bad enough, the fracturing of society's epistemic norms is catastrophic.").

[245] For a review of the canonical position about the dangers of veracity tests in U.S. constitutional law, see Leslie Gielow Jacobs, *Freedom of Speech and Regulation of Fake News*, 70 AM. J. COMP. L. i278, i283 (2022) ("[P]ermitting the government to punish speakers simply because they utter a false statement would present dangers of government overreach that the free speech guarantee guards against. These dangers include anointing the government as a type of 'truth police,' empowering the government to engage in selective prosecution of speakers with viewpoints critical of it or that it otherwise dislikes and chilling the expression of protected speech.").

[246] For an analysis of the constitutionality of these laws, and a review of similar legislation proposed in the United States, see Ayesha Rasheed,

A similar attempt had been undertaken by Brazilian President Jair Bolsonaro, who issued an executive order to give platforms' users "the right not to have user-generated content excluded, suspended or blocked," except for a limited number of "just causes" that amount to different types of illegal content.[247] The order was almost immediately rejected by the Senate.[248]

Texas and Florida's laws do not directly enforce Texans' and Floridians' community norms. Rather, both state legislatures attempt to prevent what they perceive to be "a handful of billionaires in San Francisco that run these tech companies" from acting as "the gatekeeper[s] of free speech [as] they want to be."[249] Preventing the imposition of these out-of-state community norms on communications among the states' denizens, they think, will also prevent the "silenc[ing of] conservative ideas [and] religious beliefs."[250]

This policy, however, is based on a fundamental misunderstanding of how speech regulation informs the public sphere. Texas and Florida legislatures seem to think that, in a pre-digital environment, state power was the only force preventing individuals from speaking their minds in whatever anti-social way they wanted. Since First Amendment law inhibits the state from regulating all sorts of controversial discourse (from non-obscene erotic materials to hate speech),

---

Dormant Commerce Clause Constraints on Social Media Regulation, 25 YALE J.L. & TECH. SPECIAL ISSUE 101 (2023).

[247] Medida provisória [Provisional Measure] No. 1068 (Sept. 6, 2021), Diario Oficial da União [Union's Official Gazette] Sept. 6, 2021, Edição Extra, §§8.A.VI (translation from Artur Pericles Lima Monteiro, *Unofficial English Translation of Brazilian Provisional Measure 1068, on Content Moderation* (Sept. 13, 2021), https://ssrn.com/abstract=3922992 [https://perma.cc/B2VH-U6KV].

[248] *See* Jack Nicas, *Bolsonaro's Ban on Removing Social Media Posts Is Overturned in Brazil*, N.Y. TIMES (Sept. 15, 2021), https://www.nytimes.com/2021/09/15/world/americas/brazil-bolsonaro-social-media-ban.html [https://perma.cc/6AHS-6QME].

[249] Shawn Mulcahy, *Gov. Greg Abbott Backs Bill to Stop Social Media Companies from Banning Texans for Political Views*, TEXAS TRIB. (Mar. 5, 2021, 8:00 PM), https://www.texastribune.org/2021/03/05/texas-greg-abbott-press-conference-social-media [https://perma.cc/688S-D8NC] (quoting Texas Republican State Senator Bryan Hughes).

[250] *Id.*(quoting Texas Governor Greg Abbott). It is quite clear that the governor sees Texas as the relevant political community here.

then it follows that all legal speech must be equally valuable in the marketplace of ideas. Digital platforms—the reasoning must continue—came to disrupt this free space, by creating a private forum in which private power prohibits discourse that the state purposely allows. In order to stop private platforms from censoring valuable speech, the legislatures assume it is justified to impose on platforms a requirement to provide users with something similar to First Amendment protections.

But, as we explained above,[251] this conception is profoundly misguided. The fact that the state chooses not to prohibit some types of expression directly does not turn free speech into libertine speech. In the pre-digital environment, speech was still socially regulated by a variety of gatekeepers, who evolved to be aligned with societal values. In this way, communities regulated their own public sphere with no need for formal command-and-control ordinances. By analogizing platforms to state power, and therefore removing platforms' capacity to regulate speech, Florida and Texas have not brought the Internet closer to what the public sphere was before. Rather, if their attempt succeeds, they will have converted platform discourse into an unprecedented realm of completely free-floating discourse, in which only the most nasty and brutish speech is heard.

***

These three examples show how states' attempts to reverse the bypass effect come at a serious price, one that has probably not been adequately considered. In the first two cases (direct enforcement of civility norms and the fight against fake news) the state decided to replace the void left by civil society with governmental power. By doing this, states ignore Böckenförde's paradox, attempting to secure by themselves the preconditions of democratic coexistence. As Böckenförde could have told them, this enterprise risks falling into illiberalism, preventing social norms from evolving organically and depriving the public of any means of resistance against the

[251] *See supra* Part III.

social norms the regime finds useful or desirable.

The strategies deployed by Florida, Texas, or Brazil fall into the opposite vice. They find that platforms are disrupting their community by imposing outside speech norms, and attempt to prevent this by subjecting them to the same limitations that apply to state action. This step, however, is likely to promote anarchy and not stability. The function previously exercised by civil-society gatekeepers has found no new owner. If allowed to pursue these policies, Florida and Texas will develop a completely unregulated public sphere in which public debate will become not free but impossible.

By desperately reacting to the perturbation brought by platforms, both types of state federalist strategies forget the crucial role of civil-society institutions in shaping and disciplining public debate. The natural path to explore, therefore, is to reconstitute the role of these institutions in platform communication. We explore this "civil society federalism" in the next Section.

B. Civil Society Federalism

Turning to state federalism as a means to counteract the bypass effect restores some lost gatekeeping functions to local actors, but also transitions these roles onto government shoulders. This shift teeters dangerously on the brink of illiberalism.[252] Of course, some state officials are sagacious enough to prevent this potential transgression from fully materializing, but their wisdom does not negate the fact that they are setting up a framework that could be misused by less scrupulous authorities down the line.

The inclination of state federalism toward illiberalism is not a bug, but rather an inherent feature. This is not a malfunction or an unexpected deviation—it is the design. In our rush to combat the bypass effect, we might just be sowing the seeds of a new problem.

A different path is possible. Rather than hastily replacing civil-society institutions as guardians of the public sphere, we suggest a legal plan aiming to reinstate the influence of local civil society within the intricate landscape of digital communication. Given the absence of technological barriers

[252] *See supra* Part III.

that once naturally positioned civil-society institutions as gatekeepers of mass communication, restoring their power would demand deliberate regulation requiring platforms to give domestic civil society a substantial role in curating and structuring public discourse. We call this strategy *civil society federalism.*

We envision civil society federalism as a policy agenda that is sensitive to Böckenförde's dictum. On one hand, it actively attempts to enable the processes that maintain domestic community norms alive. On the other hand, it seeks to avoid the risks of illiberalism that stem from involving the state in speech regulation in direct and insidious ways. Should these policies prove successful, they have the potential to reinvigorate the delicate equilibrium that has long characterized liberal democratic societies.

If we want to recover the virtuous equilibrium lost in the passage from the pre-digital public sphere to the platform-dominated public sphere, we need civil-society institutions to regain their gatekeeping power. Such a restoration will not happen spontaneously. In the pre-digital age, civil-society institutions had the means to block information from spreading if it did not pass their thresholds. Their tests could refer to civility (avoiding certain forms of hate speech or otherwise offensive utterances) or to epistemology (choosing only trusted sources, double-checking information, or assessing its public relevance).[253] As we discuss above, civil-society institutions power over the public sphere was not derived from laws and regulations: their control over the gate was sufficient to ensure the functioning of the public sphere.

In our present platform-dominated era, the public's dependency on civil-society institutions for communication has significantly diminished. Massive, instantaneous communication can transpire without requiring the mediation of civil society. However, should we aim for these civil -society institutions to reclaim their capacity, we need to reinsert them

[253] *See supra* Part II.

into the information flow, making them indispensable again, and not so easily bypassed.

Calls for enhancing civil-society organizations' role in platform governance are not new. Hannah Bloch Wehba, for instance, has suggested that "[p]latforms could . . . explore more robust avenues for participation by civil society in formulating and expounding policies and rules."[254] In Section VI.C.1, we delve into potential pathways to involve civil society organizations, with an aim to shape public discourse. Our propositions, while far from all-encompassing, serve to illuminate possible roles for these organizations within the context of civil society federalism. In Section VI.C.2, we elaborate on the methodologies and standards that should steer the selection process for determining which civil society organizations are given the privilege to partake in the governance of digital platforms.

C. Civil Society Federalism in Practice

1. Content Flagging

One potential sphere for the application of civil society federalism is as a second layer of content moderation through flagging of content in violation of community norms. Social-media platforms frequently influence public dialogue by suppressing diverse expressions that conflict with their internal regulations, typically referred to as "community standards."[255] While most of the enforcement is effected automatically,[256] some troubling statements slip past these automated systems, paving the way for the crucial role of users. Any member of the

---

[254] Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 76 (2019).

[255] *See generally* Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526 (2022) (exploring the complexities of content moderation governance systems). Content moderation is not the only system of rules that social networks apply, they also comply with local laws––which is a part of what we call state federalism. *See supra* note 226 and accompanying text.

[256] *See* Daria Dergacheva, *Platforms Overwhelmingly Use Automated Content Moderation, First DSA Transparency Reports Show*, LAB PLATFORM GOVERNANCE, MEDIA & TECH. (PGMT) (Nov. 8, 2023), https://platform-governance.org/2023/platforms-overwhelmingly-use-automated-content-moderation-first-dsa-transparency-reports-show [https://perma.cc/2RSU-BNGL].

digital community can flag content they deem to be at odds with a platform's community standards.[257] Be it hate speech or misinformation, users have the power to call attention to such instances, empowering the platform to deliberate over whether the content should be flagged or suppressed. Users, in effect, serve as critical partners in shaping the conversational landscape on these platforms.

We propose civil-society institutions acquire a preeminent role in flagging content for moderation. This is not an entirely new idea: social-media platforms have already started to partner with civil-society institutions and other private actors for them to become "trusted flaggers" of content. While the precise role of these "trusted flaggers" remains quite opaque in self-regulatory arrangements,[258] lawmakers have explored this practice to regulate the digital public sphere in some jurisdictions.

The European Union's Digital Services Act (DSA), passed in 2022, adopts this strategy.[259] It requires very large platforms to give priority in content-moderation review to "trusted flaggers," which is a status awarded by a governmental agency. Under the DSA, Member States award "trusted flagger status . . . to entities, and not individuals, that have demonstrated, among other things, that they have particular expertise and competence in tackling illegal content and that they work in a diligent, accurate and objective manner."[260] The Regulation provides that these entities should be limited and "can be public in nature . . . or they can be non-governmental

---

[257] Kate Crawford & Tarleton Gillespie, *What is a Flag for? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 NEW MEDIA & SOC'Y 410, 411 (2014).

[258] Naomi Appelman & Paddy Leerssen, *On "Trusted" Flaggers*, 24 YALE J.L. & TECH. 452, 453-57 (2022).

[259] Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act), O.J. (L 365) 1, 2022 [hereinafter *Digital Services Act*].

[260] *Id.* ¶ 61.

organisations and private or semi-public bodies."[261] In particular, the Regulation encourages "industry associations representing their members' interests" to apply for trusted flagger status.[262] In a 2018 recommendation—prior to the enactment of the DSA—the European Commission advised that trusted flaggers "have the necessary expertise and carry out their activities as trusted flaggers in a diligent and objective manner, based on respect for the values on which the Union is founded."[263]

The European trusted flagger policy might be a sensible one, but it is not tailored to reestablish domestic civil-society influence over the public sphere, nor to reestablish the preeminence of local speech norms. This is especially evident in the way the DSA regulates the selection of trusted flaggers. First, it does not have an explicit requirement that the institutions selected be a part of any particular civil society.[264] While it is likely that this will happen in many cases because of the expertise requirements and voluntary nature of the task, it is still likely for private companies to further their interests through trusted flagging status, or for global organizations to apply. Second, it delegates the appointment and removal of trusted flaggers fully to the government,[265] while not providing for any procedural safeguard to ensure that the state will not privilege institutions that align with their ideological commitments or electoral base.

From our viewpoint, it is essential for a trusted flagger system to have a key goal of acknowledging and incorporating

---

[261] *Id.*

[262] *Id.*

[263] Commission Recommendation (EU) 2018/334 of Mar. 1, 2018 on Measures to Effectively Tackle Illegal Content Online, O.J. (L 63/50) ¶ 5.27.

[264] *Digital Services Act*, *supra* note 259, at art. 22.2 (requiring the entity meet only these conditions: "(a) it has particular expertise and competence for the purposes of detecting, identifying and notifying illegal content; (b) it is independent from any provider of online platforms; (c) it carries out its activities for the purposes of submitting notices diligently, accurately and objectively").

[265] *Id.* ("The status of 'trusted flagger' under this Regulation shall be awarded, upon application by any entity, by the Digital Services Coordinator of the Member State in which the applicant is established . . . .").

local speech norms in content-moderation procedures. Since it is not practical to distill these norms into clear-cut decision rules, achieving this goal necessitates direct involvement from local civil-society organizations in the moderation process.[266] In response, regulations around trusted flagger designations should mandate this status to be principally, if not wholly, granted to local civil-society institutions. As civil society federalism strives to mitigate the risk of power tilting alarmingly towards the state, it is a sensible principle that trusted flagger status should largely be granted to non-governmental organizations.

2. Establishing Content-Moderation Rules

While the practice of trusted flagging can indeed be advantageous for the digital public sphere, its role, it must be said, remains somewhat circumscribed. Users can highlight specific infractions of the community standards, yet their influence doesn't extend to the crafting of those standards. They might very well be able to direct the spotlight, but they still lack a seat at the drafting table.

Content moderation rules, moreover, are crafted and implemented in extremely opaque ways, and are usually drafted at a highly abstract level. Although in some cases general community standards do prohibit specific expressions (such as blackface or Holocaust denial[267]), most of them are

---

[266] The reason why civil society organizations, and not the government, should be involved in this process is not that the former possess a superior understanding of civility norms. Rather, it arises from the desire to protect free speech by avoiding government regulation. Civil Society and traditional media are also a better choice than private market entities, since, at least rhetorically, they operate with public-facing missions, which lends them greater credibility in matters of speech regulation.

[267] *See* Jacob Kastrenakes, *Facebook Bans Blackface and Anti-Semitic Stereotypes in Hate Speech Update*, VERGE (Aug. 11, 2020, 5:08 PM), https://www.theverge.com/2020/8/11/21363890/facebook-blackface-antisemitic-stereotypes-ban-misinformation [https://perma.cc/MSL6-QY9N].

drafted in terms general enough as to be compatible with a wide array of cultural understandings. For instance, Facebook "prohibit[s] the usage of slurs that are used to attack people on the basis of their protected characteristics" such as "race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease."[268] However, the line that demarcates objective description with a slur,[269] or even the acceptable or tolerable uses of slurs,[270] are extremely context- and culture-dependent. It is simply impossible to provide enforcement rules at a general level. It is also impossible to create such rules without applying particular culture-specific social standards. Accordingly, platforms have extremely specific guidelines for their (algorithmic and human) content moderators, which often remain secret.[271]

Under our proposal, civil-society institutions have a preeminent role in establishing the specifications for the implementation of these abstract community standards. Local civil-society institutions are better placed than global platforms themselves to assess the acceptable boundaries of civil speech, the contours of tolerable humor, and the acidity of satire.

To achieve the goals of civil society federalism, platforms must devise independent operational rules for each significant

---

[268] *Hate Speech*, TRANSPARENCY CENTER – FACEBOOK, https://transparency.fb.com/policies/community-standards/hate-speech [https://perma.cc/4CMD-WCL7].

[269] The different use cases (some hateful, some not) of the n-word is case in point. *See, e.g.*, Gregory S. Parks & Shayne E. Jones, *A Critical Race Realist Analysis of the n-Word within Hate Crimes Law*, 98 J. CRIM. L. & CRIMINOLOGY 1305, 1310-13 (2007). *See also supra* note 89 and accompanying text.

[270] *See, e.g.*, Matal v. Tam, 582 U.S. 218, 246-47 (2017) (holding that the government cannot deny trademarks for potentially offensive or derogatory terms, thereby allowing the use of slurs in trademarks under the First Amendment's free speech protections).

[271] For reporting on leaks of such documents, see Andrew Liptak, *Leaked Moderation Guidelines Reveal How Facebook Approaches Handling Graphic Content*, THE VERGE (May 21, 2017, 3:19 PM), https://www.theverge.com/2017/5/21/15672242/facebook-files-moderation-guidelines-graphic-content-online [https://perma.cc/4FC6-7M7H]; Angela Chen, *A Leaked Excerpt of TikTok Moderation Rules Shows How Political Content Gets Buried*, MIT TECH. REV. (Nov. 25, 2019), https://www.technologyreview.com/2019/11/25/102440/tiktok-content-moderation-politics-protest-netzpolitik [https://perma.cc/C837-XEJ2].

political unit, and include civil-society stakeholders in the formulation of these rules. In the current arrangements, social-media platforms "train" civil-society organizations to act as their "trusted flaggers;"[272] we propose it should be the other way around.

Note that some social-media platforms have already started to seek external assistance to better craft their community standards or define their implementation in controversial cases. However, they have sought to do it through recourse to the legitimizing veneer of law, rather than civil society. Take Meta's "Oversight Board," sometimes referred to as "Facebook's Supreme Court."[273] The Board is a quasi-judicial body that revises specific content-moderation decisions made by Facebook and Instagram in a written and reasoned way. The Board is also routinely asked to provide policy recommendations to ensure that community standards are applied in accordance with freedom of expression and other human rights.[274]

Meta's Oversight Board, however, is designed in a way that privileges abstract, legal legitimation over the "local knowledge" of civil society organizations. While the Board is aspirationally composed of individuals coming from diverse geographical backgrounds, there is no rule providing for cases to be solved exclusively by members coming from the same

---

[272] *Assessment of the Code of Conduct on Hate Speech Online. State of Play*, INFORMATION NOTE FOR THE JHA COUNCIL 12522/19, 4 (Sept. 27, 2019) ("Since the signature of the Code, Facebook/Instagram have organised a total of 51 training sessions on its community standards in relation to hate speech, for up to 130 civil society organisations operating as trusted flaggers.").

[273] *See generally* Evelyn Douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. TECH. 1 (2019) (explaining the functioning of the Oversight Board).

[274] *See generally* Brenda Dvoskin, *Expert Governance of Online Speech*, 64 HARV. J. INT'L L.J. 85 (2023) (critically exploring the relationship between the Oversight Board's opinions and international human rights law).

country or region.[275] While they conduct research about the local context in which a certain expression was uttered, there is no intention nor legitimacy for the Board to intervene in the social norms that constitute discourse in the corresponding community.[276] While they accept comments by third parties prior to solving a case, there is no mechanism to privilege or encourage comments coming from the same community.[277]

Rather than exclusively relying on a specialized entity, or working in conjunction with it, digital platforms should actively seek the expertise and input of civil-society organizations when making these consequential adjudications.[278]

Through recruiting civil-society institutions for content moderation enforcement and regulation, social-media platforms could reinforce community norms within a local public sphere, partially reverse platforms' tendency towards untamed universalism, and mitigate the effects of the bypass effect.

3. Common Trusted Sources

The crisis of misinformation caused by the bypass effect threatens society.[279] Historically, mass media fostered political community by creating shared experiences.[280] Pre-digital civil society gatekeepers excluded content violating their norms while amplifying content aligned with their worldviews. Their

---

[275] There is, however, a rule that at least one of the board members assigned to a case needs to be from the region of the case. *See* OVERSIGHT BOARD BYLAWS, OVERSIGHT BOARD r 3.1.3 (Feb. 2023), https://oversightboard.com/attachment/578723797462534 [https://perma.cc/MD28-V278].

[276] *See* Abiri & Guidi, *supra* note 12, at 114-22.

[277] *See Rulebook for Case Review and Policy Guidance*, OVERSIGHT BOARD 9 (Nov. 2020), https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance [https://perma.cc/73CR-VEWP] ("The Board may request public comment briefs. Calls for public comment briefs will be posted on the Board's website with requirements for form and substance, and a deadline for submission. Submissions will be shared with the panel.").

[278] Hannah Bloch-Wehba has also suggested a version of this mission for civil society organizations. *See* Bloch-Wehba, *supra* note 254, at 76 ("Avenues for participation in adjudication could also be broadened: platforms could appoint an 'amicus curiae' to help make more informed and robust decisions in difficult cases regarding content or privacy.").

[279] *See supra* Part III.

[280] *See supra* Section II.B.

media control enabled them to curate common informational experiences, establishing common ground among citizens.

However, this shared foundation is at risk due to the bypass effect, coupled with the fragmentation of media brought on by ad-targeting and engagement-focused recommendation algorithms. To counteract this, we propose the regulation of these very algorithms. Such regulation would strive to recreate an environment in which political community members encounter common information sources, ones carefully curated by the guiding hand of civil-society elites.

Governments, for example, could require digital platforms within their jurisdiction to prioritize a transparent set of domestic "trusted sources" in their feeds.[281] For instance, during the COVID-19 pandemic, platforms voluntarily increased the visibility of information from the WHO and the CDC.[282] Such measures should extend beyond major crises, fostering common media experiences among citizens, akin to

---

[281] A variation on this idea has been around for a while. *See* Sandra Baron & Rebecca Crootof, *Fighting Fake News Workshop Report*, YALE INFO. SOC'Y PROJECT 8, https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf [https://perma.cc/MN23-7HYJ] ("Some favored developing 'whitelists' of articles or news sources, based either on user or an independent institution's ratings.").

[282] Kang-Xing Jin, *Keeping People Safe and Informed About the Coronavirus*, META (Dec. 18, 2020), https://about.fb.com/news/2020/12/coronavirus [https://perma.cc/MZ6K-D9XC] ("We've launched these globally over the last few weeks in all languages on Facebook, directing people to the WHO. In several countries we are directing people to their local ministry of health. For example, in the US we are directing people to information from the Centers for Disease Control and Prevention [CDC] and in Singapore, we're directing people to the Singapore Ministry of Health."). There is evidence that the adoption of anti-COVID misinformation by digital platforms had significant effects. *See, e.g.*, Traci Hong et al., *Effects of #Coronavirus Content Moderation on Misinformation and Anti-Asian Hate on Instagram*, 00 NEW MEDIA & SOC'Y, no. 0, 1, 1 (2023) (showing that content moderation on social media platforms, such as Instagram, has been effective in reducing misinformation and anti-Asian hate related to the COVID-19 pandemic).

traditional mass media. Such steps may not instantly solve the misinformation crisis, but they will, at the very least, prevent the epistemic divide from widening further.

It is important, however, that governments are not overly involved in the process of selecting and validating trusted sources, which according to critics can create "government-regulated 'whitelisted' media [which] often becomes a proxy for state-sponsored or government-approved news."[283] As with the other instances of Böckenförde's paradox that we have seen in this Article, by taking steps to create a common media experience the state would be risking illiberalism. Governments, therefore, should not determine or revise the list of "trusted sources." Rather, regional platform chapters should handle this task through discussions with stakeholders and the public, including civil society institutions.

Governments, instead, can mandate platforms to create transparent, publicly accessible lists of trusted sources, representing diverse political viewpoints. By giving priority to these sources in users' feeds, platforms will increase the probability of a common media experience among users within a particular polity. Citizens and organizations should have the ability to appeal the list of trusted sources, providing a feedback mechanism to improve the quality and representation over time. Furthermore, the list should be re-evaluated and updated every few years through a transparent process, ensuring it keeps pace with changes in the media landscape.

There are two main possibilities for how platforms could create trusted sources lists that represent diverse viewpoints. First, governments could demand that platforms operating in their territory work together to create one common, inclusive list of trusted news sources that they all share. This would likely require compromise between platforms to arrive at a list that satisfies government requirements for diversity of viewpoint. Second, governments could mandate that each platform follow a defined procedure, involving consultation with relevant stakeholders, to develop their own customized list of trusted sources. Either approach aims to produce a reasonably inclusive list of trusted sources across the political spectrum that algorithms would then recommend. For instance, in the

[283] Baron & Crootof, *supra* note 281, at 8.

United States, such lists would likely need to incorporate major conservative outlets like Fox News alongside other mainstream sources. An inclusive common list, though imperfect, could help mitigate the current situation where networks like Fox and CNBC depict two completely different realities. If these outlets become part of a shared media diet for the citizenry, this may facilitate more common understanding and debate of issues. However, prioritizing trusted sources would not entail granting them complete hegemony over feeds. Platforms should retain flexibility to maintain some diversity in users' feeds based on their individual preferences and choices. This equilibrium between a common media experience needed for public discourse and individual customization enabled by the internet may be optimal.

D. Equity and the New Civil Society

Our policy proposals are directed to artificially reconstruct the role of civil-society institutions in shaping the public sphere and taming the bypass effect. However, there is a crucial difference between the pre-digital civil society and the one we are proposing to empower. Before platforms, civil-society institutions chose themselves through organic selection.[284] Any attempt to legally reconstitute civil-society influence would necessarily require an intentional selection procedure. The conundrum we face is intricate, for those entrusted with the task of selection will consequently wield massive power over the public sphere. Therefore, figuring out the appropriate methodology for choosing civil-society organizations to

[284] Ruud Koopmans, *Movements and Media: Selection Processes and Evolutionary Dynamics in the Public Sphere*, 33 THEORY & SOC'Y 367, 375 (2004) ("Together, the notion of a bounded communicative space that can only accommodate a small minority of the variegated candidates for entry, on the one hand, and the selection mechanisms of visibility, resonance, and legitimacy, on the other, form the basic building blocks of an evolutionary model that allows us to explain why some actors and some messages flourish and others perish in the competition for the scarce resources of public attention and legitimacy.").

reassume a gatekeeping role becomes an essential question within the framework of civil society federalism.

The pre-digital era communicative environment, while evocative of simpler times, was not neutral. Civil-society gatekeepers were never a mirror of society's diverse and multifaceted population. Elite gatekeepers left many voices unheard.[285] The price countries paid for having elites governing their public sphere was to give them disproportionate, if inevitable, power over the civility norms that were enforced. In this process, many minorities saw their own norms being disrespected or ignored.

While judging the past is something that should be done with humility, the same does not apply to thinking about the future.[286] While we maintain that public discourse depends of agreeing on a certain set of community norms, as we shape the selection procedure in civil society federalism, we stand before an opportunity to sculpt the character of the new elites tasked with that mission.[287] It is a chance to envision a more equitable landscape, ensuring diverse, representative, and principled voices lead our public discourse, fostering a more balanced and inclusive society.

Selecting civil-society organizations to serve as new gatekeepers in the context of civil society federalism requires a transparent, reasoned, and impartial procedure. It is crucial to create a robust selection framework, ensuring that the entities chosen are autonomous, reputable, experienced, public-facing,

---

[285] *See generally* HERMAN & CHOMSKY *supra* note 162, at xii (arguing that mass media reflect the interests of the elite and that therefore "views that challenge fundamental premises or suggest that the observed modes of exercise of state power are based on systemic factors will be excluded from the mass media even when elite controversy over tactics rages fiercely").

[286] Democracy entails the endless possibility of imagining alternative futures that we can build collectively. *See* ROBERTO MANGABEIRA UNGER, WHAT SHOULD LEGAL ANALYSIS BECOME? 20 (1996) ("A democratic experimentalist will not stand waiting for the next magical moment. Rather than have us be crowned by history, he will insist that we crown ourselves").

[287] *See* MURRAY MILNER JR., ELITES: A GENERAL MODEL 7, 9 (2015) (endorsing the insight coming from elite theory that "[t]he emergence of elites is a virtual inevitability in any large organization or complex society." Even in the case of revolutions or *coups d'état* that overthrow the regime, "a society will soon be ruled by elites largely committed to their own interests").

and representative of significant societal voices. The precise form of this selection procedure will depend heavily on the domestic context. However, we can suggest some general principles.

The selection procedure should combine state and platform power—probably through platform application of state-mandated rules. The state could issue general rules as to how to select civil-society organizations to ensure a balanced and professional public sphere. However, in order for governments not to manipulate the selection, it is safer for the procedure to be implemented by the platforms themselves.[288] In this crucial point, our proposal departs from the DSA, which entrusts governments with the appointment of trusted flaggers.[289]

The procedure should establish clear selection criteria. First, to ensure the impartiality of these gatekeepers, they should be independent of government and political parties, which would protect them from potential influence or manipulation. Second, these organizations must have a proven track record of experience in managing public discourse and a strong reputation in society, demonstrating their credibility and expertise. Third, the organizations should have a public-facing mission—they should be committed to benefiting the public. This does not automatically exclude all for-profit organizations, but their public mission should be paramount. Although the precise profile of organizations will be heavily dependent on the national context, universities, think tanks, NGOs and even mass media are likely candidates to fulfill this role.

---

[288] In some federal arrangements, the federal state sets general goals and constituent units are entrusted with securing the means to achieve them. The best-known examples of this mechanism are the European Union's "directives," which "shall be binding, as to the result to be achieved, upon each Member State to which [they are] addressed, but shall leave to the national authorities the choice of form and methods." *See* Consolidated Version of the Treaty on the Functioning of the European Union, art, 288, May 9, 2008, 2008 O.J.

[289] *See supra* notes 258-262 and accompanying text.

While the selection criteria are important, the design of the procedure itself is also relevant for the outcome and its legitimacy. The selection procedure should begin with an open call to all organizations that meet the relevant criteria. To maintain transparency and trust in the process, the selection should be publicly announced, detailing the purpose, criteria, and process involved. This openness serves the value of transparency, but also of diversity. By making the call open, platforms can offset their own biases and ensure that a wide array of organizations, channeling diverse voices within society, can apply.

Moreover, the procedure should follow basic principles of procedural justice, to ensure legitimacy and maximize the chances of correctness.[290] Thus, the selection process should be reasoned and provide grounds for decision-making. When an organization is selected or rejected, the reasoning behind that decision should be communicated clearly.[291] There should also be a mechanism in place for organizations to appeal the decision. This enhances fairness and legitimacy, allowing for any overlooked considerations to be rectified, and providing an outlet for organizations to be heard again after initial rejection.[292] This system could be supplemented by external oversight of the selection procedure—conducted by an external entity like Meta's Oversight Board or an ad-hoc reviewer—offering an additional layer of checks and balances.

---

[290] Multiple empirical studies have shown that fairness in procedures enhance the legitimacy of regulations and compliance thereof. *See, e.g.*, TOM R. TYLER, WHY PEOPLE COOPERATE: THE ROLE OF SOCIAL MOTIVATIONS 14 (2010) (expounding a behavioral model of according to which "[p]rocedural justice and motive-based trust lead to favorable dispositions and, through them, motivate voluntary cooperation on behalf of groups").

[291] Glen Staszewski, *Reason-Giving and Accountability*, 93 MINN. L. REV. 1253, 1278-94 (2009) (examining the reasons why "reason-giving fosters democratic legitimacy").

[292] On the functions of appeal to enhance procedural legitimacy see generally MARTIN SHAPIRO, COURTS: A COMPARATIVE POLITICAL ANALYSIS 49 (1986) ("One of the functions of a 'right of appeal' may be to provide a psychological outlet and a social cover for the loser at trial.").

Conclusion

The platform Internet is not splitting, at least not all the way. Instead, we are witnessing the messy birth of a federal power sharing arrangement between global digital platforms and nation-states. While we hold that power-sharing is preferable to localism and universalism, the realization of a stable and mutually beneficial platform federation remains a distant aspiration.

In this Article, we propose a reinvigoration of the gatekeeping role of civil-society institutions, allowing them to shape the domestic norms of public discourse once again. This way, we seek to safeguard the civility and respect inherent in public discourse, without falling into the pitfall of undue state influence or global dominance.

That path, alas, will be a bumpy one. Academics have been warning about the decay of civil-society institutions for decades.[293] Social media is routinely accused of accelerating this decline in social cohesion,[294] and—if the most apocalyptic among the critics are right—perhaps it will be the last nail in democracy's coffin.[295] Social media might, indeed, be creating a society of unapologetic individualists, narcissistic teenagers, and apathetic idiots. We are in a situation where the detrimental influence of social media and the gradual erosion of civil-society institutions are inextricably linked. To address this, we need to re-energize these institutions, bolstering their role in defining the parameters of public discourse.

Even if civil society is truly in decline, we have no option

---

[293] *See generally* PUTNAM, *supra* note 136 (surveying the decline of social capital in 20th Century America).

[294] For a mixed review, see Sandra González-Bailón & Yphtach Lelkes, *Do Social Media Undermine Social Cohesion? A Critical Review*, 17 SOC. ISSUES & POL'Y REV. 155 (2023).

[295] This is an all-too-familiar prediction. *See e.g.*, Jonathan Haidt, *Yes, Social Media Really Is Undermining Democracy*, ATLANTIC (July 28, 2022), https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975 [https://perma.cc/M28G-SFSN].

but to resist. Liberal democracy as we know it is facing momentous challenges, but we work under the obligation of keeping it alive–at least until we find an acceptable alternative. We must make a similar commitment with regards to civil society and social-media platforms. It might be the case that civil-society organizations are doomed to gradually decay and fade out from our political lives, but we should not allow that to happen without a fight. If democracy and communication is increasingly happening in platforms, then civil society must be brought onto platforms.