

## 6

### Platform Democracy Now!

We should democratize platforms. Ultimately, this means that the leaders of companies, as well as of governments, will have to give up some control. But I contend that doing so will be in their own best interests – for companies, their profit interests; for governments, their interests in protecting the safety of their people and the integrity of their own institutions.

#### 6.1 CAN PLATFORMS BE GOVERNED DEMOCRATICALLY WITHOUT STATES?

If it wasn't clear before one of the most important social media platforms was taken over by an erratic billionaire with a penchant to make decisions out of spite, it certainly became clear thereafter that platforms suffer from a severe democratic deficit. But can they be governed democratically without subjecting them to the direct supervision of nation-states? I contend that the answer is yes.

The concept of a “democratic deficit” comes from the literature on the European Union (e.g., among many others, Follesdal and Hix 2006; Sánchez-Cuenca 2017). I use it in this context as a way of capitalizing on the distinction that the EU context invites us to offer between “governed by democratic nation-states” and “governed democratically.” The EU could suffer from a democratic deficit even though it is made up of nations that are themselves democratic (and hence even though the people of member nations exercise ultimate control over its decisions) because its own institutions may not give its people adequate effective power over the decisions that it makes that affect their lives, for example, because there are too many administrative layers between the exercise of democratic political autonomy and government policies.<sup>1</sup> But democracy and governments can come apart in the opposite

<sup>1</sup> As of the last few years, it is no longer true that the E.U.'s member states are democratic: Hungary, Poland, and Turkey (the latter a transitional member) are more reasonably described as at least semi-autocratic. Still, the broader point holds: The democratic deficit conversation long precedes the turn to autocracy of those countries.

direction too: An entity may be governed democratically through means other than by being under the direct control of a democratically ruled nation-state. Thus, we may speak coherently of “workplace democracy” (Landemore and Ferreras 2016), and, indeed, some scholars have suggested that the solutions to the EU’s democratic deficit might come not through the empowerment of national parliaments but through the integration of civil society organizations into its framework for exercising authority (Greenwood 2007; Steffek, Kissling, and Nanz 2008; Kohler-Koch 2010). Such suggestions are not without substantial controversy; it may be that the only morally valuable forms of democracy go through nation-states. But the question is at least open.

If this is true, then there’s conceptual space for it to be true both that (a) existing (liberal) democratic nation-states lack the capacity to effectively regulate behavior on platforms, at least not without sacrificing other important liberal-democratic values, and (b) platforms, and behavior on platforms, might still be subjected to democratic control in ways that do not directly draw on nation-state power. And identifying this possibility might point the way to identifying the real regulatory gap in platforms: It may not be that we want to turn their actions, and the actions of people using them, over to the direct control of nation-states, but that we want those paired set of actions to be regulated *democratically*. This could be carried out by, for example, subjecting platform rules and enforcement processes – or perhaps even their broader scope of affordances – to the control of those affected by them.

I also wish to resist the supposition that “democratic” requires any specific institutional form of public interaction.<sup>2</sup> As with every other persistent problem in democratic theory, Dewey (1927) is instructive: He argued for a functionalist conception of democracy not attached to any particular processes. On his account, there’s an inherent conflict of interest in the exercise of public power, because of course that power is exercised by officeholders with their own interests, thus, “[w]hen the public adopts special measures to see to it that the conflict is minimized and that the representative function overrides the private one, political institutions are termed representative” (Dewey 1927, 76–77). I believe we can adopt this idea to fill out the concept of a democratic deficit. Platforms do not suffer under a democratic deficit because there isn’t enough voting, or because existing liberal democracies do not exercise enough direct control over them. Rather, they suffer under a democratic deficit because those who operate them are insufficiently constrained to act in the interests of the public at large, as defined by the public at large.

In the context of global platforms with global publics, this suggests that the real goal is to render platform power accountable to the entire world. No small task. But this chapter sketches out one way in which we might achieve it.

<sup>2</sup> Dewey (1927, 144–145) explains that representative institutions that emerged from one set of social circumstances may be unsuited for different circumstances.

6.1.1 *Against Naive Participation*

Starting in 2009 and running for about three years, Facebook recognized, in at least a dim sense, its democratic deficit, and tried to achieve something like electoral accountability (or at least sham legitimation) via holding platform-wide votes about its rules. Even by its own terms, the process was a miserable failure. In 2009, the company held an election that managed to recruit the participation of less than 1 percent of its users. That vote did not even count, because the decision rules Facebook established for the system required 30 percent of the company's users to participate before the result would bind the company. One might have thought that in a massive international platform where users have even less incentive to vote than in domestic elections, a 30 percent participation threshold is so unrealistic as to be irrational (or, more cynically, rigged to never generate a binding outcome) (see discussion in Bloch-Wehba 2019, 75). Finally giving up on the idea, in 2012, the company held an election to abolish elections (Robertson 2018). A huge majority of the participants in that final, doomed, election voted against the Facebook proposal to abolish their own vote (Engelmann, Grossklags, and Papakyriakopoulos 2018, 96–97). But, once again, turnout fell below the 30 percent threshold, so their objections were ignored, and Facebook democracy fell with a resounding whimper.

There are lots of failures in the process which we might blame for that ridiculous waste of the time of all involved. The elections were partly characterized by administrative incompetence – some users reported, for example, that Facebook's email notifications of the votes landed in spam, including in a spam-like “other messages” folder on Facebook's own messaging system (Engelmann, Grossklags, and Papakyriakopoulos 2018, 101). Surely a sizeable chunk of the blame is due to the fundamental unseriousness of the enterprise with its 30 percent threshold.<sup>3</sup> But some of the blame is also doubtless due to the ordinary pathologies of elections. National elections, with much higher stakes and much smaller electorates, often feature low turnout, and political scientists have long debated the rationality of voting, given the effort cost involved and the low individual benefits (Dowding 2005). It should be unsurprising, then, that the effort to evaluate complex policy changes was not worth many users' time – one LA Times columnist estimated that figuring out what they were voting on in the first election would require “an hour of eye-strain” (Sarno 2009).

Even if a Facebook election had managed to reach that 30 percent threshold, what could such a vote have meant? A simple up or down majority vote on complex multidimensional policy documents across hundreds of millions (then) of highly diverse users from wildly different cultural backgrounds seems, at the least, poorly suited to make

<sup>3</sup> It could be worse. Shortly after his acquisition of Twitter, Elon Musk held a poll through his own account on whether to reinstate Donald Trump (<https://twitter.com/elonmusk/status/1593767953706921985>).

meaningful decisions. Formally speaking, this seems like a recipe for Arrovian electoral difficulties, such as cycles (see Riker 1982, 119–23); more abstractly, it is not obvious why every user in a multinational platform ought to be ruled by a majority vote when it might be possible to both make democratic decisions on individual policy changes separately and make them differently for different subpopulations of users. I am reminded of Dewey's reaction to the sudden complexity of industrial democratic politics:

It is not that there is no public, no large body of persons having a common interest in the consequences of social transactions. There is too much public, a public too diffused and scattered and too intricate in composition. And there are too many publics, for conjoint actions which have indirect, serious, and enduring consequences are multitudinous beyond comparison, and each one of them crosses the others and generates its own group of persons especially affected with little to hold these different publics together in an integrated whole.<sup>4</sup>

This sort of development had thwarted the nineteenth-century politics which America had inherited when Dewey was writing; and, in a similar sense, it thwarts a kind of simple version of democratic inclusion like that which Facebook attempted (or pretended) to implement. For there isn't likely to be a majority interest or a kind of Rousseauian common good amenable to simple majoritarian decision, as opposed to a multiplicity of overlapping and conflicting interests; moreover, the incentives for individual participation are vanishingly small on a platform the size of Facebook. More carefully designed institutions are needed to adapt to the nature of the underlying public.

For those reasons, this chapter describes much more institutionally complex recommendations.<sup>5</sup> It describes a model for participatory entities to be created for the platform ecosystem as a whole. In the first instance, this chapter focuses on social media platforms, although with some modifications, the design it sketches could be expanded to transactional platforms too.

Although such entities could be created by the companies themselves – and doing so ultimately would be in their best interests – I recognize that they might be unable or unwilling to do so alone. Accordingly, I also argue for some steps that governments should take in order to facilitate the development of those institutions (as well as other discrete reforms).

Ultimately, the core recommendation of this chapter overlaps with, though extends on, extant proposals for “social media councils” by some scholars and advocacy organizations.<sup>6</sup> To the extent of their overlap, this book may be seen as, in part,

<sup>4</sup> Dewey (1927, 137).

<sup>5</sup> With complexity, naturally, comes contingency: The more detailed a proposal, the less likely it could be implemented without significant modification in light of the real-world understanding of stakeholders.

<sup>6</sup> Article 19 (2021); Article 19 (2019); Docquir (2019); Article 19 (2018); Tworek (2019); Stanford Global Digital Policy Incubator (2019); Tworek et al. (2020); Kettemann and Fertmann (2021).

providing a social scientific framework for understanding the likely effectiveness of such councils as well as a normative framework for understanding their value.

In addition, however, I aim to go beyond existing social media council proposals in terms of scope and ambition. The Article 19 proposal, which is the most developed (and most frequently cited in the literature), essentially amounts to a version of the Oversight Board, but with more democratically accountable selection processes, potentially also combined with a forum for more widely participatory discussion of the relationship between social media content moderation and international human rights standards, and potentially also organized into multiple regional forms to deploy local knowledge and values. While such councils would be an immense advance over existing governance institutions, they nonetheless could go further by (a) expanding their ambitions to a full integration of the public into rulemaking processes; (b) setting a foundation for their later expansion beyond social media into platforms more generally; and (c) more fully addressing opportunities and incentives for those presently not well represented by governments or dominant civil society organizations to participate. This chapter attempts to push further in these directions.

The other prominent extant proposal along lines similar to those of this chapter is in a working paper by Aviv Ovadya on what he calls “platform democracy.”<sup>7</sup> Ovadya identifies many of the self-control problems rooted in pressure from politicians that I described in Chapter 5, and argues, as I do, that offloading sensitive decisions to the general public can mitigate those problems. He argues, accordingly, for the creation of “citizen assemblies” to make such decisions. From that perspective, this chapter can be read as providing a kind of constitutional framework for such assemblies.

Before we get to the proposals, however, it will be useful to reiterate and summarize some of the design criteria entailed by previous chapters.

## 6.2 WHAT WOULD EFFECTIVE DEMOCRATIC PLATFORM GOVERNANCE LOOK LIKE?

The prior chapters entail a number of design criteria, both in terms of inputs (what should successful governance institutions look like) and outputs (what kinds of behaviors and outcomes should be generated). This section sketches some of them to set the stage for a more detailed articulation of the proposed system.

### 6.2.1 *Overlapping Governance Entities at Multiple Scales*

A key lesson from the research on polycentric governance is that multi-scale entities can be effective in permitting the incorporation of interests that themselves

<sup>7</sup> Aviv Ovadya, Platform Democracy, working paper at the Harvard Kennedy School Belfer Center for Science and International Affairs, October 18, 2021, [www.belfercenter.org/publication/towards-platform-democracy-policymaking-beyond-corporate-ceos-and-partisan-pressure](https://www.belfercenter.org/publication/towards-platform-democracy-policymaking-beyond-corporate-ceos-and-partisan-pressure).

naturally exist at different or multiple scales. Jurisdictional overlap is not to be feared, as this permits negotiation between different compositions of stakeholders, innovation through learning by individuals involved in multiple governance groups, and potentially some degree of competition. Furthermore, if natural compositions of stakeholders across different scales and locations do not provide sufficient diversity, artificial groupings can be constructed of different kinds of stakeholders (in the platform context, users, nonusers, workers, civil society, and governments, as well as different ethnic, religious, and national groups).<sup>8</sup> In terms of the typology of multistakeholder structures articulated by Raymond and DeNardis (2015, 580), the design most likely to achieve the benefits associated with the resource governance and democratic knowledge literature is what they describe as “heterogeneous polyarchy.”

A key goal of such an organizational structure is to permit informational cross-pollination between center and periphery. The structure should encourage people on the periphery to use their local knowledge by giving them actual power to influence platform decisions. Moreover, the center, in other words corporate personnel, should be in a position to provide feedback to individuals and groups on the periphery, so that the latter may learn how to more effectively use the power they have. This is particularly important in the platform context, where public understanding of the affordances and consequences of platforms, as well as of the realistic potential for altering their operations and the consequences of those operations, has been hampered both by the nontransparent organization of the companies and by the technical nature of some of the most significant tools available to them (such as machine learning).

With overlapping jurisdictional governance entities, as well as in recognition of the fact that complex social and political issues will come into play in the most important decisions, conflict is inevitable. One supplemental design principle applying to such conflict is that there should be a built-in preference for negotiation over adjudication – or at least for forms of adjudication that are inherently negotiation based, such as by having higher-level “appellate” style decision-making bodies that are composed of representatives of the lower-level decision makers whose decisions are in conflict.

In addition to the notion of polycentricity, another concept associated with Elinor Ostrom and the Bloomington School is a highly influential list of eight (or so, depending on how one counts sub-items) conditions empirically associated with successful resource governance. While not all are relevant for the current project, several are of particular importance here. Of the conditions as described in a 2010 article (Ostrom 2010a, 653), particular emphasis for the present project is due to:

<sup>8</sup> Drawing on the Athenian example discussed in Chapter 3 (see Ober 2008; Traill 1975): Groups of otherwise socially distant individuals could, for example, be randomly subdivided into Cleisthenes-style platform *phylai* according to a stratified sampling process that ensures the mixing of people from a variety of social and jurisdictional groupings; such *phylai* could then be allocated specific governance responsibilities.

3. Collective Choice Arrangements: Most individuals affected by a resource regime are authorized to participate in making and modifying its rules.
  - 4A. Monitoring Users: Individuals who are accountable to or are the users monitor the appropriation and provision levels of the users.
  - 4B. Monitoring the Resource: Individuals who are accountable to or are the users monitor the condition of the resource.
6. Conflict Resolution Mechanisms: Rapid, low cost, local arenas exist for resolving conflicts among users or with officials.
7. Minimal Recognition of Rights: The rights of local users to make their own rules are recognized by the government.
8. Nested Enterprises: When a common-pool resource is closely connected to a larger social-ecological system, governance activities are organized in multiple nested layers.

Of those, 6 and 8 are essentially the description of the structural features of polycentricity just given; 7 requires some modification, as we are here concerned with recognition by governments as well as companies; while 3 and 4 are descriptions of the powers that governing groups should have. I take this to be a core description of what a polycentric governance arrangement for platforms should look like.

### 6.2.2 *Genuine Empowerment of Workers, Users, and Nonusers*

There are at least three constituencies that can serve as countervailing forces to the power of both platform executives (Chapter 4) and governments (Chapter 2), as well as contribute knowledge to executives and governments (Chapter 3) – if they have enough power to give them an incentive to do so. Users are an obvious group – those who use platforms currently lack a meaningful voice in their governance, particularly in geographic areas and languages in which those users lack social connections to core company personnel. Likewise, nonusers are an obvious group – with respect to platform externalities, nonusers in many cases have as strong a stake in governance as users, and may have a distinctive perspective on the harms caused by platforms in part by virtue of likely being distinct from users in systematic fashions (such as by occupying underserved geographic, linguistic, or occupational groups); the knowledge of nonusers for that reason is also likely to be valuable in governance projects.

Platform company workers are worthy of particular attention. Of course, many of those whom platform companies consider their workers – core employees in engineering functions in company headquarters, for example – are substantially more capable of influencing company outcomes than in the nontechnology industries, partly because of the longstanding culture of relatively decentralized autonomy within Silicon Valley companies, and partly simply because such workers have relatively higher bargaining power than in most other industries (though this power

varies depending on the current state of the employment market in technology). However, which workers have that power and in which companies is more a feature of internal corporate politics than any robust design choice – for example, employees in functions most relevant to governance can easily find themselves sidelined if opposing constituencies have a more central role in company politics.<sup>9</sup> Incorporating those workers into governance institutions – particularly those which incorporate the general public – is likely to primarily serve constraint-related functions in line with the argument of Chapter 4: Such workers are likely to have access to knowledge that external constituencies might find useful to sanction companies for violating their commitments and also likely to have at least some capacity to have their hands on levers to influence outcomes directly. Thus, putting such workers into interaction with the general public (and civil society) can help empower the public.

More importantly, however, platform companies have many workers with effectively no participatory role in company decisions even as they have superior knowledge – and a claim to themselves be a relevant constituency whose interests are at stake. I refer, of course, to the vast legions of content moderators. Many of those workers are employed by subcontractors, often though not always offshore, and in notoriously oppressive and quasi-Taylorist work conditions (Roberts 2019; Newton 2019; Chen 2014). The offshore character of many of these workers is particularly relevant not only to the injustice of their treatment but also to their actual capacity to improve platform governance.

A particularly striking example comes from the Philippines. According to Adrian Chen's 2014 expose of the working conditions of Facebook's content moderators, many of those workers were in the Philippines. Yet only two years after Chen's article, the Philippines became a site of one of the most notorious failures of Facebook's content moderation, a failure that apparently took the company by surprise, as Rodrigo Duterte weaponized the platform to menace his political opponents. According to a *Washington Post* article from several years later – reporting that the political troll problem in the Philippines had gotten so bad that malicious actors there were offering their services as specialists in other countries – “After apologizing for not acting sooner, the company has staffed up a local office in Manila – a rare move for Facebook – and launched a digital literacy program” (Mahtani and Cabato 2019).

The problem should be clear. *Facebook already had a local office in the country* – or, at least, if Chen's article is to be believed, the company already had lots of workers there, it's just that those workers were supervised by a subcontractor and had no way to communicate their own knowledge of political conditions in the Philippines

<sup>9</sup> This is, for example, one standard story about what happened with Facebook (e.g., Horwitz and Seetharaman 2020). Similar claims have been made about Google (e.g., Harwell and Tiku 2020; Tiku 2020).



to Menlo Park, nor sufficient influence over platform outcomes to give them any incentive to do so even if they had somehow had the means.<sup>10</sup> This example suggests that directly incorporating content moderators – who have both local knowledge in the countries they occupy as well as platform-specific knowledge about the kinds of things that are happening at any given time within a platform – into governance structures is likely to be particularly beneficial. And, of course, it may also contribute to relieving the subordinated status of those workers.

It bears repeating that genuine empowerment implies incentive-compatibility: Users, nonusers, and workers of various stripes should have an incentive to invest their time and energy in platform governance. This can (and should) mean paying them, but it also means, as discussed in Chapter 3, giving them a clear way to influence outcomes, and hence motivating them to act when they care about those outcomes.

### 6.2.3 *Iterative Stakeholder Inclusion and Experimentation*

One inevitable fact about the identification of stakeholders, and of the appropriate representatives for stakeholders, in negotiating and decision-making processes is that any process for doing so will be flawed. In every real-world form of political decision-making, both the identification of the rightful stakeholders with a legitimate participatory interest and the representatives or other processes for taking into account the views of those stakeholders are permanently contested. Thus, the history of every political community that has ever been even partly self-governing has featured intense conflict over membership and exclusion – what rights would the Metics have in Athens, or the Socii in Rome? How about immigrants in the United States?

Platforms face the same problem in a slightly different form. Suppose you agree that they ought to reduce their democratic deficit by formally incorporating participation in their governance decisions from members of the public. And suppose you think that those members of the public who participate ought to do so organizationally. Now we have to decide which organizations get to participate. This is difficult enough in the wealthy countries in which platform shareholders and employees tend to live. But consider foreign countries, particularly the countries that have been the victims of colonialism, and who have nondemocratic governments. The governments of those countries cannot represent the interests of those who themselves are illegitimately excluded (imagine the government of Myanmar claiming that any involvement it wants to have in a platform's decision-making is sufficient to represent the interests of the Rohingya). But for any set of, for example, civil society

<sup>10</sup> As far as I can find, I'm the first person to point this out – could this really be true? – which suggests that the purely bureaucratic difference between core company employees and offshored contractors has become reified to the point that it's difficult to even think of the possibility of those two groups of people serving similar functions.

organizations (or even groups of people to be selected from via sortition) who might be given a seat at the table, there is inevitably and always some other set of organizations that might reasonably be argued to better represent the relevant stakeholders in a given community. And the representation decision – to decide which organizations count – is itself legitimated only by fair representation. What a mess!

The analogy to the boundary problem in political states is clear. In each case, the paradox is that we need to decide who legitimately gets included in order to legitimate the decision about who gets included (in political states, cf. Whelan 1983, 15–16, 40–41). Only a democracy can fairly decide its boundaries for things like suffrage, but a state that hasn't already democratically decided its boundaries can't bound the population who gets a vote; only a platform that already fairly represents the interests of the Rohingya can fairly figure out whom to invite to fairly represent the interests of the Rohingya.

The solution is to give up the fantasy that it might be possible to define the boundaries of who is to be included *ex ante*. Rather, one must build institutions in such a fashion that they are capable of responding to unexpected claims for inclusion, with a bias toward responding in the affirmative to such claims. (Recall James Baldwin's insight from the previous chapter: This is how the collective figures out its identity.) This is of a piece with the general adaptive character of effective forms of governance in complex organizations, discussed in Chapter 3, along with the Deweyian experimentalism discussed in the Introduction. That is, platforms and overarching governing organizations must accept that they can never be perfectly representative, and will inevitably be subject to critique for imperfectly legitimate and imperfectly rights-respecting decisions. However, they can be responsive to such critiques by engaging in a perpetual process of revision as to the composition and organizational design of governance mechanisms in order to improve their representative character over time.

That being said, at a bare minimum any design worth considering even as a starting point must be reasonably inclusive. It must, for example, incorporate representatives from countries in the Global South, as well as those who are currently underrepresented by existing states, such as indigenous peoples. And it must have some mechanism for minimizing the risk of repressive cooptation by authoritarian governments while still including the people of the states under the thumbs of such governments. I shall characterize any such arrangement as “democratic” as a reflection of the idea that the core of democracy is the exercise of authority by groups of ordinary people, determined as the output of a good faith effort to include all of those who have a legitimate stake in the decisions to be made, and chosen in a fashion that is not corrupted by the powerful (as opposed to any particular process of selecting those people, such as election).

An approach of design and membership experimentalism is also likely to be necessary to make any novel governing institutions realistic. The circumstances of politics are such that the creation of any such institutions will inevitably not go

according to plan, not only because of the demands for inclusion that groups can be expected to make but also because of the unanticipated needs and claims of governments, civil society organizations, and companies themselves. Thus, any framework for governance is necessarily tentative, even hesitant.<sup>11</sup>

#### 6.2.4 Robustness, of a Sort

In *Democracy and Knowledge*, Ober (2008, 13) offers the following criterion for judging a social equilibrium (i.e., social institutions):

Because social cooperation produces economic value (as well as being valuable in non-material ways), more cooperative and (in changing environments) more dynamically adaptive equilibria perform relatively well in economic terms. Less cooperative and inflexible equilibria perform poorly. An equilibrium may be judged robust if it is capable of maintaining coherence in the face of substantial environmental changes.

While the focus on economic value is perhaps too specific, the point remains if we simply focus on the abstract idea of gains from cooperation: The ultimate goal of the enterprise of platform governance is to maximize robust gains from cooperation among users (and companies), whether that is mutually beneficial economic transactions in a marketplace like Amazon or Airbnb, or communicative goods on social media. And doing so means designing institutional forms that can be resilient not only to endogenous shocks – that is, that can control the pathological forms of the novelty that emerges from diversity and scale, as discussed in the first few chapters – but also from exogenous shocks that radically change the incentives and constraints under which participants operate. A prominent example from the platform governance world would be the sudden introduction of hostile nation-state intelligence services such as Russia’s into the political discourse on social media of numerous nations. Ober’s notion of “dynamically adaptive” institutions tracks the insight, also familiar from the adaptive management context, that one way to achieve resilience or robustness (which I take to be equivalent concepts) is to build a capacity for internal modification (or learning) into an institution (e.g., Berkes 2002, 295).

However, it would not necessarily be appropriate to directly import the notion of resilience as used in related literatures. The idea of resilience (or robustness)

<sup>11</sup> Some successful methods of decentralized governance have begun, not with institutional design, but with devolution, with the (apparent) idea being to circumvent centralized bureaucratic resistance but with the (I speculate) likely consequence that handing power to local groups, however imperfectly implemented, will in part be helpful to achieve the required buy-in for decentralized institutions as well as the benefits of local knowledge as to efficacious design (Isaac and Heller 2003, 82). Some degree of devolution of government power from regulatory agencies as part of a negotiated design process is likely to be a practical approach to the development of institutions like those described in this chapter as well. Then the groups to whom power has been devolved must be responsive to claims for expansion and inclusion.

as it is often used, for example in engineering, refers specifically to a return to a stable equilibrium (Ruhl 2011, 1375–78). For Ober, for example, that equilibrium arguably would be defined in terms of the system of power-sharing within Athens (and in particular its democratic character), as well as advantageous properties of its economy and international role. But it's not obvious what such an equilibrium would refer to in the context of platforms or why (beginning from their current harm-generating state), such an equilibrium would be desirable. Indeed, it's arguably inconsistent with the Deweyian learning approach taken in this book, in which governance institutions are taken to embody a process of learning, which applies not only to the domain which they govern but also to their own functioning, stakeholders, and the like – such an approach implies that sometimes *disequilibrium* is the goal, in order to facilitate change on the basis of successful learning. It may be that a somewhat more appropriate conception of resilience is “ecological resilience” in which there is “a high capacity for swings in behavior in response to changing conditions without altering the system’s basic structure and process design” (Ruhl 2011, 1381). But even there, the “basic structure” of a governance system, and indeed of platforms themselves, should be in question if the argument of this book is to be believed.

However, many of the capabilities of resilient systems are applicable to governance design even under such circumstances. There are some exogenously induced changes to which such a system should be resistant – increases in workload, attacks from hostile third parties intent on undermining the system (such as Russian spies, 4chan trolls attempting to take over a platform, or repressive governments). Moreover, the related concept of adaptive capacity (Ruhl 2011, 1388–93) which refers to the resources available within a system to choose among available equilibria or find new equilibria in the context of change, may be more relevant. Accordingly, such a system should feature the capacity to absorb new information, “response diversity” (Ruhl 2011, 1401) to generate a variety of possible responses, and numerous communicative linkages between scales of authority (Ruhl 2011, 1401), in order to maximize the chance of effective knowledge transformation.

### 6.2.5 *Robustness against Bottom-Up as well as Top-Down Threats*

Not all negative externalities arise from hostile state actors, the capture of corporations by conflicted interests, or even the clueless governance choices of heterosexual, cisgender, white males in wealthy countries. Any account of how ordinary people ought to be included in a governance enterprise ought to acknowledge that those who are disregarded, subordinated, and oppressed themselves are also capable of generating emergent harms at scale. Including, for example, ordinary people in India in platform governance doesn't merely mean including people who might push back both against the assumptions and interests of Menlo Park and Modi but also people who themselves might be inclined to, say, Hindu

nationalism. Disregarding that fact itself amounts to a nasty kind of colonialism, which imagines the other as a kind of “noble savage” or an innocent, free of the pathologies of politics.

To some extent, the answer to this problem rests on the Churchillian cliché that democracy is “the worst form of government except for all the others” – that however poor the record of governance-from-below in making decisions rooted in error and malice, the record of governance-from-above is immeasurably worse. However, some democratic designs might be better at defending against threats from below than others.

One lesson that might be borrowed from federal forms of government such as that in the United States is that multilevel forms of bottom-up governance, properly designed, can limit the impact of the harms generated by inclusion. Polycentric federal systems, in which different levels of governance not only have defined “territories” (physical or virtual) but also are brought into interaction with one another within and across levels to solve shared problems also incorporate a built-in checking capacity in which communities with their own internal autonomy become accountable to others as equals. Such arrangements, at least potentially, can mitigate the risk of threats from below while still capturing the governance benefits of inclusion.

### 6.3 DEMOCRACY WITHIN A SINGLE PLATFORM: A SKETCH FOR REDDIT

The focus of this chapter is platform democracy *across* platforms. However, it is important to note that this chapter is also compatible with the notion of platform democracy within a single platform. I sketch an example in this section to illustrate its potential, particularly in the context of polycentric defense against bottom-up threats.

One common model for social media platforms is what we might call a place of places, or community of communities. The quintessential example is Reddit, but other prominent examples include Discord and elements of Facebook (particularly groups). On this model, the central platform provides some subset of network infrastructure and core user experience, search/discovery/recommendation, and baseline rules and enforcement (governing both subcommunities and individual users), while individual subcommunities provide an additional layer of rules and membership selection directed to their own goals.<sup>12</sup> In a way, this platform model harks back to the early days of the Internet in the form of things like chatrooms, BBSs, usenet groups, and the like – but with the difference that the contemporary social media platforms are likely to be far more widely used and to provide far more sophisticated recommendation algorithms to drive users to subcommunities.

<sup>12</sup> I am unavoidably reminded of Robert Nozick’s (1974) minimal state “framework for utopia” in the form of subcommunities.

Such platforms pose distinctive governance challenges: On the one hand, they facilitate (potentially beneficial) diversity and a market for good governance, insofar as they permit individual self-selection into groups with goals and governance models most consistent with their own interests and goals. On the other hand, such groups also create numerous social externalities. The distinctive subgroup-based externalities can broadly be lumped into two categories: (a) radicalization, in which group members reinforce dangerous false beliefs and recruit others into those beliefs (consider the example of “blackpill” and “incel” subreddits which reinforce reactionary and misogynist beliefs among young men) and (b) coordination, in which such groups are used to organize harmful actions against others.<sup>13</sup> Such groups can also harm their own members, for example by leading them into unwise decisions (consider anti-vaccination groups). There is evidence, for example, that Facebook’s “groups” surface became host to immense amounts of election-related misinformation circulating among Trump supporters in the days leading to the January 6 insurrection (Silverman et al. 2022; Timberg, Dwoskin, and Albergotti 2021).

For the sake of concreteness, we might think of two specific prominent reddit groups: r/TheDonald, a reddit group composed of fans of Donald Trump, which was banned for hate speech in 2020 (not to be confused with a different Trump fan subreddit which was banned after the January 6 insurrection (Isaac and Conger 2021)), and r/WallStreetBets, which briefly exploded in membership and influence when it became the epicenter of a mania over stock in moribund companies like (most famously) GameStop (Harwell 2021). Those two subreddits are core examples of governance problems because both were at least nominally legitimate in inception (unlike more unproblematically malicious or even criminal examples, such as r/TheFappening, devoted to sharing stolen nudes of female celebrities), yet arguably caused substantial amounts of external harm.<sup>14</sup>

<sup>13</sup> The border between these two categories of externality can blur, as when radicalization groups also encourage antisocial behavior by individuals within them, or when antisocial behavior is actually carried out through communication on those groups (such as swapping of nonconsensual sexual imagery of third parties).

<sup>14</sup> One question we might ask is whether such groups ought to be any more governed than intermediate associations in the physical world. After all, people get radicalized in real world associations, but liberal democracies recognize this risk and tolerate it for the sake of associational freedom; the same goes for groups that pose a threat of coordinated actions to harm others: Until they actually do something, we typically do not suppress them on the basis of such a risk or even talk (before r/DonaldTrump there was the Ku Klux Klan). Thus, for example, the crime of conspiracy traditionally requires some “overt act” to distinguish actual criminal enterprise from mere talk. However, a platform intervening on a Facebook group, subreddit, or the like is different from the state intervening on ordinary associations in the physical world, not merely because platforms have less power than states and can only break up one particular communication channel for some group, but also because groups may pose greater dangers on platforms for at least two reasons. First, recommender algorithms offer groups additional recruiting affordances – and one way platforms might choose to govern groups is by withholding the benefit of those algorithms (Gillespie 2022) – by contrast, the state doesn’t recommend the Ku Klux Klan (these days). Second, because of the pervasiveness of social media platforms (e.g., on smartphones), it’s possible for a vulnerable individual to be lured into far deeper involvement in an

The structure of communities-of-communities platforms makes them potentially ideally suited for experimentation with a polycentric governance strategy. The key idea is that these groups share discursive space (users' recommender-algorithm-created feeds as well as society at large) and hence arrangements that permit groups to impinge on one another's use of that space will likely lead to it being used in ways that are more compatible with its shared nature.

Consider, for example, what might occur if Reddit classified all very large subreddits (by some reasonable metric, such as "joined" membership, daily active users, and daily pageviews) as "high-impact subreddits" and implemented a platform-wide rule that both content rules and content-moderation decisions by moderators of those subreddits are subject to appeal to a randomly selected committee of moderators of other high-impact subreddits, which may overturn those decisions by a supermajority rule. To have an impact on subreddits like *r/TheDonald* where harmful externalities are caused by neglecting to act rather than by acting, of course, the failure to moderate content might also be subject to appeal.

Such a first-pass proposal would be subject to one serious problem: Depending on moderator capacity, this system would be ineffective because such a task would be incredibly burdensome for any such committee. But less burdensome strategies might work better. Suppose that instead (following the "fewer traffic cops, more speed bumps" strategy suggested by Sahar Massachi) Reddit conferred on a committee of external moderators the power to impose group-level restrictions on a subreddit, such as rate limits for posts or distribution limits. Perhaps such a committee would not need to identify a pre-existing rule violation to impose such limits, but rather would act based on the conclusion, justified by a reasoned written opinion (itself potentially subject to review either by the company or by ordinary members of the public on something like the terms described below), that existing activity on a subreddit is socially harmful. The social harm justifications for limitations in the written opinions, in turn, could be incorporated as rules into platform policies in a subsequent adjudicative process in order to achieve something like the regulative effect that Chapter 5 attributed to the Meta Oversight Board. This directly captures the core idea of polycentric governance, that decisions about the exploitation of a shared resource (in this case, a broader discursive environment) should be made in a way that brings competing users into interaction and negotiation at varying scales.

There are two obvious objections to such a system. The first is that it would permit arbitrary decisions – the imagined committee would not need to identify a rule

online group relative to a physical group – one's family and close friends are more likely to notice and intervene if one spends multiple hours every day going to Klan meetings relative to spending those same hours on a racist subreddit. Moreover, private entities can, and hopefully do, refuse services that dangerous groups might rely on to associate and communicate in the physical world as well. One imagines that many landlords would, for example, refuse to rent meeting space to their local Klan chapter. Similarly, I like to think that many universities and similar physical-world intermediate associations would resist efforts to create something like a Klan student group.

violation in order to restrict a subreddit. The second is that it would tend to suppress the degree of diversity on any such platform – we might expect outlier and unpopular groups to be subject to more restrictions than more mainstream groups.

The first objection tracks a longstanding dilemma of political governance: Any highly diverse society that generates lots of behavioral novelty will also sometimes generate emergent threats – serious dangers to important shared goods that, because of their novelty, could not have been anticipated by rulemakers in advance. In political states, the ideal of the rule of law is typically taken to restrict the capacity of states to address such emergencies, at least with coercive force. Permitting these kinds of rule-less restrictions is in tension with this value. In response to this objection, I would simply note that subreddits already contain at least two sites of arbitrary power, in individual subreddit moderators as well as in Reddit as a whole. By creating a method by which multiple subreddit moderators working together could intervene in the operation of individual subreddits by overcoming nontrivial collective action hurdles, this section probably sketches a system that increases, rather than decreases, the overall degree of constraint of power in the system as a whole.

However, to the extent there is a lingering rule of law worry in such a system, it can be alleviated in two respects. First is by connecting it to the broader system to be described in a few pages, which includes an appellate process for content moderation decisions. To the extent within-platform democratization was plugged into across-platform democratization, such as by granting cross-platform councils appellate power over the sorts of rate limits and other restrictions imposed by within-platform democratic groups, that would build in further protections against arbitrary exercises of power. Second is by mechanical tweaks to the method of exercising the power of actors like external moderators on Reddit, with an eye to establishing a workable balance between the capacity to respond to emergent rogue behavior and the goal of preserving the principle of rule-bound regulation. For example, a history of committees acting arbitrarily in cases of relatively low-grade necessity may counsel creating or increasing a supermajority requirement before a committee may act, or imposing a second-order rate limit on the number of times committees can act, or some other abstract restraint increasing the cost (including the opportunity cost) of imposing a burden on some subreddit.

This same strategy can be used to alleviate the homogenizing effect identified by the second objection. Again, there is an evident balance to be struck: More diverse platforms also mean more dangerous externalities to society at large, so the degree of homogenization can be tweaked over time and in an experimentalist spirit by increasing or decreasing the ease of intervening on a rogue group. In addition, the advantage of composite decision-making entities like a large committee operating by a supermajority requirement is that the composition of such committees can also be weighted in the interest of diversity – ensuring, for example, that all political perspectives, social, ethnic, and national groups, and so forth extant on the platform are



significantly represented reduces the likelihood that actions will be taken solely on the basis of the strangeness of a group. Moreover, on a community-of-communities platform, groups that are subject to restraints from others can even be overrepresented on subsequent committees (i.e., giving the Donald Trump fans a slightly higher probability of serving on other committees in the future) in order that they may exercise countervailing pressure on the discursive environment as a whole.<sup>15</sup> Such a mechanism would also potentially impose a healthy bias on the committee system toward limited actions to address short-term problems rather than more pervasive sanctions against disfavored groups, as being subject to those more pervasive sanctions would give sanctioned groups more power in the rest of the committee decisions down the line.

While the remaining proposals in this chapter focus on a kind of constitutional structure for polycentric participatory governance spanning many platforms, the foregoing section suggests that there is also substantial scope for polycentric participatory governance within individual platforms. Further development of individual-platform strategies is beyond the scope of this book, but such strategies could be integrated into the multi-platform system and are to be encouraged.

I now return to the global constitutional structure for platforms.

#### 6.4 A DEMOCRATIC CONSTITUTION FOR GLOBAL SOCIAL MEDIA

The design criteria articulated in this chapter can be potentially achieved with a rather dramatically more aggressive version of the social media council concept advanced by the scholars and advocacy groups cited above.

Accordingly, I now sketch a system of robust, inclusive, and democratic social media councils. As the previous sentence suggests, the model below is crafted for the specific case of social media platforms, and would likely require some modification to also apply to transactional platforms. However, there is nothing particularly limited to social media platforms in the general arc of the model. While this model is quite ambitious, I argue in the next section that it is in fact readily doable. However, it should be noted that the details given below are *merely* a model, meant to illustrate a design for a system of participatory involvement in social media governance that can satisfy the design criteria given above; any concrete implementation would undoubtedly require modification (and democratic input).

As a first pass, I note that this system as envisioned would be industry-wide, that is, there would be a common set of participatory institutions across social media companies in the first instance and other companies subsequently.<sup>16</sup> An ideal form of initiation would have a legal framework for these councils (e.g., in the form of the

<sup>15</sup> Back to Aristotle, ultimately, and the idea of ruling and being ruled in turn.

<sup>16</sup> Compare extant speculation about the possibility of expanding the ambit of the Meta Oversight Board to serve multiple companies (e.g., Hatmaker 2021).

appropriate exemptions from antitrust law, trade secret law, and the like) established by international treaty, and the initial participants would be the largest multinational social media companies (primarily Meta, Alphabet, and Twitter; perhaps also Reddit, TikTok – to the extent compatible with its Chinese government involvement – and other companies with moderately large userbases), with the option to be joined, as desired, by smaller social media companies as well as nonsocial media platform companies such as Amazon and Airbnb to the extent compatible with their business needs.

In order to preserve the diversity of the ecosystem of platforms with distinct identities in the course of an overall system of governance, on matters concerning a particular company, the membership of the councils below should be supplemented by a significant minority of representatives specifically associated with (though not selected by) the company involved. This could include, for example, workers and users who spend a particularly large amount of time on the platform in question.<sup>17</sup> Such individuals could bring a sense of the discrete identity of a particular platform to a deliberation. The framework for worker representation described in the Conclusion can also serve as a tool for finding company-specific representatives who are nonetheless not controlled by companies.

The system centers on three layers of councils: A *local layer* composed of a diverse ecosystem of councils with members selected by sortition among both users and nonusers of any specific platform;<sup>18</sup> a *regional layer* composed of members selected by local layer councils; and a *global layer* composed of members selected by regional layer councils.

The number of local councils is meant to be large, such that it is in principle possible for there to be one council per nation-state, plus councils corresponding to major nonstate groups, such as indigenous nations, religious communities, national minorities, and the like; with the initial set of local councils determined by consultation among states, NGOs, companies, and international human rights personnel. Each local council should be a sufficient size to represent a meaningful sample of its underlying constituency, subject to the constraint of being sufficiently small to permit reasonable deliberation – with perhaps a cap of fifty members – selected at random with a sizeable supermajority (perhaps 80 percent) of membership composed of the group which it is meant to represent, but with a significant minority

<sup>17</sup> One imagines fashion influencers on the Instagram panel or journalists on the Twitter panel.

<sup>18</sup> Ideally, the membership of such councils should be selected from the population at large, rather than being limited to people who use any platform, however, in view of the difficulties of selecting among entire populations, the relative ease of selecting among platform userbases (since, after all, platforms can just randomize over their account holders), and the risk of intimidation associated with having governments carry out the selection in the case of autocratic governments, selection from the superset of the userbases of all platform participants (with perhaps some stratification to bias selection toward smaller platform userbases so that not all participants are users of a single or small number of dominant platforms) can serve as a reasonable proxy to start with, but with the caveat that the system must be responsive to demands for inclusion by groups of nonusers with reasonable selection processes.

randomly selected from the population at large.<sup>19</sup> The purpose of this composition is to provide a balance between the interest of creating bodies capable of exercising some power of self-determination by specific social groups, and the informational advantages of diversity within local councils. As noted above, local councils should also have company-specific sections for each company that participates in the system, and incorporating workers (including, critically, contractors) and users of those specific companies.

Each local council should be provided with staffing support sufficient to carry out the tasks described below, training (to the extent its members desire it) in deliberative processes as well as existing platform rules and systems, and, most importantly, sufficient pay for members to give them a strong incentive to participate, relative to local economic and political conditions.<sup>20</sup> Council members could be selected for six month terms, renewable (at an individual's option) once, but with some small subset of long-term members (say, five years) selected by each council for institutional continuity. Members should be permitted to serve anonymously to protect them from external pressure.

Importantly, local councils should have complete control over their own internal processes, including which of the powers described below they exercise and under what conditions, as well as decision-making rules such as whether they operate by majority vote, supermajority, consensus, or some other procedure, and the like.

Local councils should have four key powers. First, to serve as first-line content moderation adjudicators, accepting appeals from users dissatisfied with the result of platform moderation decisions. Second, to directly communicate local concerns to platform companies and international bodies, with privileged access to company policymakers, human rights organizations, and – to the extent compatible with the character of their local governments – government officials.<sup>21</sup> Third, to propose local modifications to company policies to adapt to their particular cultural circumstances. Fourth, to receive information about the impact of platform operations directly from ordinary people as well as local governments and NGOs, and, as appropriate, transmit that information through privileged channels to companies and others.

Of those, the power of proposal is most complex: I envision something like a quasi-administrative process to be carried out on the company side, where the

<sup>19</sup> This, of course, may require adaptation for the sake of language compatibility, although sufficient funding should be made available to provide translators.

<sup>20</sup> “Sufficient pay” can be measured by the acceptance rate among people selected. Training in deliberative processes as well as platform systems could resemble that of (or even be provided by) the scholars and NGOs who currently conduct training in deliberative polls (Fishkin, Luskin, and Jowell 2000), existing rule of law and similar trainings in recipient countries under the international development regime, and/or activist groups who provide training in things like consensus process and other tools for collective decision-making. However, such training should be purely optional, so as to not amount to an interference in group-specific methods of decision-making.

<sup>21</sup> In light of the important role such councils would exercise, it is of course likely that they would *de facto* also have privileged access to the press and NGOs.

submission of a proposal for a local rule variation by a council requires a reasoned response from the company to which the proposal is directed similar to the “notice and comment” rulemaking process under American administrative law, which response would be made public, and hence subject to the constraints of broader political and market mechanisms which might sanction companies. For example, interested governments might use the failure of a company to convincingly explain why some rule variation is unwarranted as a trigger to consider legislation on the subject. Of course, with inexperienced members, many proposals can be expected to be impractical, particularly with respect to difficult questions such as bounding any locally applicable policy (and local councils may choose to make joint proposals covering multiple groups). However, an important advantage of a reasoned notice and comment process with a discrete group with institutional memory such as a local council is that such discursive interaction can also help refine both council and company senses of what is possible (and ultimately educate the general public as well) in an iterative learning process that could ultimately result in practicable proposals.

The regional layer is meant to consist of a manageable intermediate number of councils, with the purpose of aggregating information from local councils, facilitating communication and coordination among them, and serving as a check on company policies.<sup>22</sup> Depending on the ultimate numbers, I envision each local council contributing between one and three members to their local regional council (selected however local councils please, from whatever population they please, with one-year terms renewable once), which should also contain representatives from workers’ councils (described in the Conclusion), NGOs, and governments that respect human rights, so long as representatives from local councils are a majority. In addition, regional councils should have some rotating representatives from local councils not in their own region in order to create beneficial cross-pollination of knowledge and ideas.

Regional councils are intended to serve three functions. First, they may decide on appeals from local councils in content moderation disputes. Moreover, it is likely that some local councils could have overlapping jurisdiction over appeals crossing multiple areas, in which case a regional council can serve as a forum for a negotiated resolution of decisional conflict.<sup>23</sup> Second, major platform user rule changes must be submitted to them (in advance, except in emergency situations) for comment – once again importing a notice and comment rulemaking process from US administrative law, but with the additional provision that a sufficiently large supermajority

<sup>22</sup> Regional councils should have the same provisions for training, staffing, payment, and decisional autonomy as local councils.

<sup>23</sup> How such a resolution would come about depends on the processes and inclinations of each council – the local councils might make independent decisions and then reconcile them with the aid of the regional council, for example, or they might agree to create an ad hoc composite council to address the situation.

of regional councils (say 80 percent) may disapprove a policy change in its entirety. Third, the regional councils may themselves submit proposals for global user rule changes to platforms, to again be addressed with a reasoned response.

Finally, the global layer is meant to consist of a single council, whose members are chosen by the regional councils and supported in a similar fashion, for two-year terms renewable once. The global council is intended to be a highly trusted participant in company decision-making, and should be given privileged access to company product designs, statistics relating to conduct violations, and the like – subject to company nondisclosure agreements (themselves subject to public interest exemptions as described in the Conclusion) as well as company objections to individual members' access.

I envision the global council to have five functions. First is to hear content moderation appeals from the regional councils, or, where appropriate and in its discretion, make referrals to platform specific bodies (like the Meta Oversight Board). Second is to suggest, subject to the same notice and comment provisions as noted with respect to the other councils, changes to basic platform affordances (as opposed to user rules). The global council is a potential source of suggestions to, for example, limit the capacity of users on a particular platform to reshare content in an effort to mitigate harmful virality. Third is to generate public reports on the scope of platform harms and company content – to, in effect, serve as a trusted third-party auditor. Fourth is to hear demands from groups of stakeholders to create their own local councils, and either approve them on its own authority or disapprove of them subject to being overturned by a majority vote of regional councils – in order to ensure that the system can, over time, expand to include a greater proportion of those who ought to be included. The final function of the global council would be to be the source of overall adaptivity, with the capacity to propose any changes to the overall council system, to be adopted on approval by a supermajority of local councils.

Such a system would give ordinary people, including ordinary people from currently unrepresented groups, a substantial amount of power to affect platform decisions – for that reason, it would give them a strong incentive to deploy their local knowledge (Chapter 3) in the enterprise. It would shift some of the balance of platform power toward the global South (Chapter 2). Being randomly selected from ordinary people (with the potential capacity to participate in secret and anonymously as needs warranted), it would be at least partly robust against the manipulation of autocratic governments – not fully robust, to be sure, but robust relative to the main alternative of giving those governments an active role in selecting participants. It would also give different groups of stakeholders numerous institutional ways to interact with one another, through the cross-group representation on the local councils as well as through the regional councils; the regional decision-making process would also likely give those groups a reason to negotiate with one another to find mutually acceptable outcomes in domains of overlapping interest or authority

(Chapter 3). Finally, the central council layer would both create a powerful check on the arbitrary authority of platform executives (Chapter 4) as well as build substantial adaptive capacity into the system as a whole to respond both to demands for inclusion of excluded groups (Chapter 2) and ensure the overall system's capacity to respond to changing environments and institutional learning (Chapter 3). In short, it is a constitutional order for the platform economy designed to respond to the scholarship reviewed in this book – a system of adaptive, polycentric participation with the capacity to transmit knowledge between center and periphery and constrain powerful company executives.

A few more words should be said on the relationship of this system to the traditional governance functions of rulemaking and adjudication. With respect to the former, this system would continue to respect the private property rights of companies by retaining their dominant position in rulemaking for their own platforms: It would require strong broad-based consensus in order to force a company to not make some rule change which it has proposed.<sup>24</sup> At the same time, it would provide a channel for direct public feedback and discussion of platform rules for the benefit of companies and the public at large, for cross-pollination of knowledge and ideas through the significant random selection and rearrangement of groupings of subcouncils, and, in extreme cases of divergence between companies and the public interest, for some degree of coercion to be wielded against companies (either directly or via public pressure) by organizations that are representative via the principle of sortition but not subject to the kinds of partisan capture to which states are unfortunately prone.

On the adjudication end, I begin by reiterating that, for the most part, the key point of adjudication isn't to provide something like individual due process in the American sense, for the simple reason that the individual stakes of individual decisions tend to be fairly low. Few people will suffer seriously from having their Facebook posts taken down. Exceptions primarily center on situations where an individual has a substantial economic stake in some platform decision.<sup>25</sup> While

<sup>24</sup> It may make sense to have a similar supermajority process in order to impose a novel rule on a company. However, this assumes a pessimistic estimate of the responsiveness of companies to less coercive and more persuasive efforts by councils, which I am unwilling to do in advance. If something like this system were adopted, one function of the global council, in collaboration with governments, would be to drive its adaptation into something with more teeth if needed.

<sup>25</sup> For example, Google is notorious for erroneously identifying people as violating their terms of service, potentially with a machine learning algorithm, and then arbitrarily canceling their accounts, with effectively no recourse or even ability to discuss the decision with a human – a decision that can be devastating if, for example, a business is built on their cloud offerings, or an individual has things like personal financial information, important intellectual property, legal documents, and so forth, stored only in Gmail, Google Drive, an Android phone, and the like. Business Insider once compared such an account termination to “having your house burn down” (Stokel-Walker 2020). Searching through the archives of sites like Hacker News and Reddit will reveal numerous allegations of seemingly arbitrary account terminations by Google as well as other major companies, such as Apple's app store. Such stories are also routinely covered in the news. For example, a user who lost access to some

these are serious problems, it seems to me fairly likely that situations of such severe economic loss are relatively rare among platform bans. Moreover, because victims of such bans have an adequate economic incentive to file suit, such cases could potentially be handled with minor modifications to ordinary domestic contract law – such as by establishing that terms of service provisions prohibiting no-process terminations, on mere suspicion, of accounts that hold customer funds are unconscionable, and requiring the return of any stored funds and the provision of access to download stored information upon termination.<sup>26</sup>

Under this design, the primary purposes of councils' adjudication function are: (1) informational, that is, to provide information to companies, to states, and to individuals about how effective current platform rules are and how consistently they are enforced; (2) precedential, that is, to provide a means for modifying rules and for accommodating emergency decisions into the rules going forward; (3) exceptional, that is, to assist companies and states in handling situations of true and severe public importance, such as the Trump ban (for example by offering advice or, if appropriately empowered, taking provisional actions like suspending dangerous accounts on their own); and of course (4) constraining, giving platforms some basis to stick to their announced rules even in the face of short-term temptation or public pressure. The council system should be able to handle those functions, and, in virtue of the number of people involved, would likely have significantly greater institutional capacity to make decisions than elite organizations like the Meta Oversight Board.

Inherently, the two sides of adjudication and rulemaking tend to expand in conceptual scope, as participatory rulemaking and enforcement trends into participatory design. By this, I mean that a number of scholars have argued that interventions on the design affordances of platforms may be a more effective

\$25,000 worth of purchased media content when Apple terminated his user account recently filed a class action suit against the company, alleging that it claims the right to terminate such accounts – and hence utterly deprive individuals of their financial investments in their devices – merely on suspicion of violating its terms of service. *Price v. Apple* case no. 5:21-cv-02846-NC (US District Court for the Northern District of California, April 20, 2021) (see Chant 2021). Occasionally, these bans seem to be imposed for downright abusive or even fraudulent reasons. For example, Amazon has reportedly banned people's accounts – including pocketing credit balances they happened to have and taking away access to Kindle books they purchased – for returning too many defective products (Brignall 2016), and in 2018, Google paid 11 million dollars to settle a class action lawsuit alleging that they terminated AdSense accounts just before payment was due (Abrams 2018). While not all of the examples just cited are about platform services as such, many of the same companies also operate platforms with equivalent economic impacts on some users – for example, Alphabet owns Google as well as YouTube, and content creators whose livelihood relies on YouTube could suffer severely from such arbitrary bans.

<sup>26</sup> In February 2021, two European professors authored a report on behalf of the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs recommending that open-ended company rights to terminate digital services be deemed unlawful (Loos and Luzak 2021, 26), so there might be some movement in this direction.

tool to control certain kinds of pathological behavior than post hoc enforcement (Massachi 2021; Kraut, Resnick, and Kiesler 2011, 6–8). For example, Goodman and Kornbluh (2021, 10) have proposed built-in friction in social media platforms to inhibit content virality, such as a “circuit breaker” automatically halting distribution of high-traffic posts to provide an opportunity for manual intervention. Hence, participatory control over enforcement at some point may shade into participatory influence over the structure of platforms more generally, for example, to implement such suggestions or to prevent the use of manipulative techniques to promote engagement at the expense of integrity. The global council concept is meant to incorporate a mechanism to facilitate this process through its capacity to propose changes to platform affordances.

The example of engagement techniques further suggests that beyond immediate platform features, the community at large as represented by users as well as governments, civil society, and so forth, are likely to want – and to demand through representative organizations like those sketched here – input on higher-level design processes and criteria, in other words on company *goals* in addition to company *tools*. The design criterion “maximize engagement” in social media is highly controversial and may also be in substantial part responsible for some of the worst pathologies of such platforms, such as the promotion of emotionally arousing but deceptive or polarizing content, and it’s entirely reasonable for stakeholders to demand a role in determining the appropriate degree and scope of its use.<sup>27</sup>

The challenge is that such decisions are the core of company product strategies, where companies are most likely to object to external intervention and even disclosure to the extent that strategies for recommendation, advertisement, and the like are also competitively valuable trade secrets. Even if it were acknowledged that external stakeholders have legitimate intervention rights in such processes, they may be impossible to implement without effectively turning councils into the companies themselves. And this may, ultimately, just be a development that companies need to embrace: Perhaps the real end state of all of this is full platform democratization.

While I confess to thinking that this would be a good idea, needless to say, companies would resist any development that would so drastically undermine the control of owners and managers over their platforms. But this simply illustrates that renegotiation would be required even of a highly developed system of public inclusion in

<sup>27</sup> No less an authority than Mark Zuckerberg came very close to admitting that the pursuit of engagement is a major part of the problem in a note on Facebook (which currently bears a “last edited” date of May 5, 2021, although it was originally written in 2018) entitled “A Blueprint for Content Governance and Enforcement” ([www.facebook.com/notes/751449002072082/](https://www.facebook.com/notes/751449002072082/)) in which he explained that for any given set of content rules, as permissible content comes closer to the “borderline” of those rules, it tends to engage more users – and (by implication) hence get more distribution. As Zuckerberg acknowledges, this also gives manipulative content creators an incentive to create such “borderline” content. (The solution he proposes is to progressively downrank borderline content.)



platform governance. That should not be surprising. I'm no James Madison (at any rate, given this book's reliance on Ober's analysis of classical Athenian institutions, perhaps a better hubristic self-comparison would be to Cleisthenes); this chapter is not meant to be a fully developed prescription for platform governance. Rather, it's meant to be an illustration of the sort of design that can meet the theoretical prescriptions described in the rest of the book.

If something like this system *were* to be implemented, modifications to the basic design could also capture different balances among the various interests at stake. For example, local or regional councils could be given the capacity to make locally binding precedents in their appeals from content moderation decisions, perhaps with overlapping jurisdictions. That power would increase the incentives for members of local councils to actively participate in adjudications, and also expand the polycentric character of the council system insofar as it would expand the range of situations where councils would be obliged to negotiate with one another to resolve conflicting decisions. However, it would also significantly increase the administrative burden of the system. In an experimentalist spirit, if any system resembling the one described in this chapter were to be adopted, it would be reasonable to choose either the precedent or the no-precedent system, and rely on the overall management role of the global council to evaluate the success of the enterprise and try out the other system in order to test the capacity for improvements. Or, perhaps, more consciously experimental methods could be deployed in which some local councils could be randomly allocated precedent-generating power as a pilot for study and possible expansion.<sup>28</sup> Similarly, if the extant system provides too much power or too little power to companies, it could be changed by, for example, giving companies a veto power over mechanisms such as the global council's capacity to modify the system as a whole.

Another possible area for variation is in the degree of information councils may require from companies. Existing evidence suggests that external governance entities may face problems in getting enough information to fully carry out their functions. In particular, Arun (2022) argues that the Trump case illustrated the weakness of the Oversight Board because the Board was denied accurate and complete information about the role of cross-check in preserving Trump's posts prior to his removal. While I have suggested that the Board as a whole nonetheless improves Facebook's governance, her points are well taken – and generate the suggestion that the global council, and possibly also the regional councils, should have the authority to compel information from companies. However, the scope of that authority and its limitations would be matters for negotiation and experimentation in any reasonable implementation of such a design.

<sup>28</sup> Balancing scientific fidelity with decolonial ambitions, councils in the global South could be over-represented among precedent-empowered councils in order to counterbalance the dominance of the United States and Europe in existing platform rules.

One obvious objection to such a system would be its expense and administrative burden. But this objection does not seem to me to be compelling. According to press reports, Facebook alone has apparently spent 13 billion dollars in a five-year period on some 40,000 content moderation personnel (Wagner 2021); one estimate has the total costs of content moderation industry-wide at almost nine billion dollars a year (Satariano and Isaac 2021). While we can expect these numbers to decline in the course of the contraction in the Internet economy as I finalize the manuscript for this book, it is almost certain that the amount of money being invested in content moderation will still be well into the billions annually. Relative to that immense enterprise, the costs of paying members and staff for a system involving a few thousand additional people are a drop in the bucket, and could be assessed to company participants proportional to the size of their userbases or their revenue through ordinary tax and transfer systems.

#### 6.4.1 *Application: Managing Cross-cultural and Intracultural Social Conflict*

To see how such a system might improve existing governance challenges, let us consider again the Thai *lèse-majesté* law. That law has been characterized as both a source and product of political repression and as a kind of religious protection similar to Islamic laws barring blasphemy against the Prophet (Streckfuss 1995; Mérieau 2019). Perhaps it's both – a bunch of Americans has no business deciding. Obviously, a system like that described in this chapter cannot answer the question whether the law is legitimate – which is ultimately a matter for social conflict and democratic process in Thailand itself to decide. But the platform constitution I sketch could potentially make it possible for platforms to come to principled and legitimate decisions about the extent to which they must participate in the law's application.

Thus, imagine that the Thai government asks a social media platform to remove some speech that allegedly defames the monarchy. And imagine that platform rules do not simply reserve the decision on legal compliance (a decision itself which requires judgment, as of course no law is wholly clear, cf. Arun 2022, 247–51) to the company. Rather let us suppose that platform rules provide that decisions on state requests for legal compliance are subject to the council system. The platform company in question might offer its own legal analysis as an input to that decision, and the rules for that decision might defer to local law, but those decisional rules might *also* take into account the overall structure and ends of platform rules (i.e., the platform's identity as described in the previous chapter). It might also take into account the character of the government making the request (on a democratic to authoritarian spectrum) to determine how much deference should be granted to local law.

Under such circumstances, the decision could be referred in the first instance or appealed to a local level council which would, if selected properly, have a

majority from the Thai people, although also with members from outside Thailand. Assuming that the selection process for that council was properly random and that the council operated under productive deliberative processes, it would have the benefit of credible internal representations of the state of Thai thought about both the interpretation of the *lèse-majesté* law and its compatibility with their conception of a legitimate government. Assuming the selection of the Thai participants was done without the control of the Thai government and the deliberations were effectively kept confidential, participants would be relatively protected from retaliation for their involvement.<sup>29</sup> That council could then make a determination, and, at its option, even issue a written opinion or written majority and dissenting opinions laying out a fair and deliberative representation of Thai public opinion on the merits and application of the law. At subsequent levels of appeal, the determination of the local council – as well as any representation of their reasoning they wish to include, and the continuing presence of local representatives on higher-level bodies – would carry forward the outcome of this deliberation, which would likely receive a substantial amount of deference from good-faith higher-level deliberators even as they introduce their own cultural perspectives, understanding of international human rights standards, and legal judgments into second-order deliberation.

In this way, a decision process for the platform implementation of the *lèse-majesté* law could incorporate local values of the community in the form of both the government's judgment (reflected in its demand under its laws) and the views of ordinary Thai people. This is analogous to a jury system in the American tradition, and we might consider a local council's refusal to implement the law if that occurs as something similar to jury nullification, with a similar democratic function.<sup>30</sup> This local judgment is then blended with more global values including those derived from platform identities (in the form of their rules and platform-specific representatives), as well as human rights and similar universal values.<sup>31</sup>

In addition to incorporating local *values*, something like the council process could also take account of relevant local *knowledge* which may be unavailable to foreigners, like about the short-term political conditions in Thailand, the extent to which the *lèse-majesté* law is currently being used for repression, and the status of the person against whom the law is being used (e.g., are they a well-known

<sup>29</sup> They cannot be completely protected – state surveillance and repressive punishment is always still possible. But no extra-state institutions can be completely protected from state repression; that would be an unrealistic goal.

<sup>30</sup> On the democratic function of jury nullification, see Gowder (2021, 53–54).

<sup>31</sup> Human rights enter the picture via the thoughtful deliberation of people of the world as a whole through the higher-level council system. However, it would also be a reasonable decision to incorporate expert bodies like the current Meta Oversight Board into the process, for example by formal requests by councils for their advice, in order to receive a more fully fleshed-out analysis of the relevant human rights principles.

dissident?). As a whole, the process described in this chapter would partly (but not completely) empower the people of Thailand to come to their own decision on the problem, and hence support their self-determination rather than platform colonialism. It would also partly (but not completely) incorporate the weight of global judgment and company interests. And it would partly (but not completely) incorporate the interests and demands of the Thai state in virtue of beginning with a process framed by their law and with a norm that their laws are to be entitled to some degree of deference. This is a compromise between the conflicting standards (moral and pragmatic) that apply to such a decision – and one which, being a compromise, will fully satisfy none of those standards (and hence fully satisfy no reader) but which would be markedly superior to the current system, and, I submit, superior to most other accessible systems which would overweight the interests either of companies or of potentially repressive states.

Of course, this is an idealization. In messier reality, the deliberations could go bad, leading to polarization or irrational decision-making. The state could use its troops and guns to manipulate or override the decision (arresting the members of the local council, banning the platform altogether). The company could do the same (defying the councils and risking whatever sanctions were implemented in the contract or the laws of states to defend their authority).

I offer two points in mitigation of the last two of those risks – that council decisions would be overridden by power, either the guns of states or the dollars of companies. First is that such overrides would be costly. A governance structure representing an inclusive compromise between a variety of competing interests is likely to recruit support among various influential publics, NGOs, governments, and the like. Under such circumstances, a company or government just overriding the decisions such a scheme generates would suffer a reputational cost at a minimum (so long as the underlying decision wasn't clearly incorrect). It is possible that they could experience more severe costs. Those costs should reduce the likelihood of such overrides occurring on a regular basis.

Second, and following from the first point, such overrides would have an important communicative function – indicating that a company or a country saw the determination as sufficiently wrong, or saw its own interests as sufficiently threatened, to incur such reputational or sanction costs. Under such circumstances, sometimes (though obviously not always) the decision *should* be overridden. This is because incurring such costs is a *credible* signal that important company or government interests are at stake. Such an override would ideally trigger a broader examination by other governments, NGOs, and democratic publics, who could help determine whether the underlying council decision failed to fully take into account the costs they were imposing on the company or on the state. In a polycentric spirit, we might represent the decision to override as a kind of forced negotiation or as an appeal to higher-order politics and markets for a final determination.

## 6.5 IS ANY OF THIS REALISTIC?

One objection to the proposals described in this chapter is a kind of apparent unrealism. Why should any company, or any government, participate in adopting the complicated polycentric governance institutions suggested by some professor? The intended audience for this book is people who work in platform companies as well as in policy roles in government rather than just academics, that is, people with actual leverage over policy. But even if anyone listens, why would we think that it's realistic to get any of these actors to create such institutional forms which involve fairly drastic compromises to their own authority?

And yet, these kinds of institutional developments are not unprecedented. An important site for parallel innovations is the European Union. Scholars have noted that the growth of central authority in EU institutions against the authority of individual states also coincided with a growth in sub-national authority centered in regions within member states, a phenomenon which went under the name "multilevel governance." Gary Marks (1996) described a number of potential motivations, meant to explain the decisions of European state leaders to promote the dispersion of power to both regional and supranational levels, which could just as well motivate company leaders to carry out a similar dispersion in the platform context. For example, he suggested that leaders might wish to tie their own hands for the purposes of negotiation with other actors, to entrench policy programs, or to shift blame for unpopular decisions to other actors (Marks 1996, 25–27). Similarly, company leaders may wish to disperse power in order to improve their capacity to negotiate with actors that might otherwise attempt to coerce their decisions (such as politicians who wish for favorable treatment of their social media content), protect their preferred policies against principal-agent problems by disaggregating power to entities that prefer those policies more than, for example, their employees do, or divert political criticism.<sup>32</sup>

We also have evidence from other governance contexts of the willingness of companies, governments, and NGOs to work together. Jessica Green describes examples in international environmental governance, such as a council composed of representatives from all three categories of entity which conducts monitoring under the 1998 Dolphin Conservation Treaty (Green 2014, 11), or the delegation of monitoring duties under the Kyoto Protocol to private companies under the supervision of an executive board established by the treaty (Green 2014, 120–30). Of course, other aspects of internet governance have long been characterized by multistakeholder arrangements, although those arrangements have been controversial due to the unbalanced power between governments and companies and in favor of US interests embedded in them (Carr 2015).

Moreover, we cannot ignore the incontrovertible fact that there is demonstrated interest in novel governance institutions, at least from the platform

<sup>32</sup> See Chapters 4 and 5 for more on these motivations.

company side. Meta did create its Oversight Board – initially at the suggestion of a law professor, no less (M. Sullivan 2019)! The various Wikipedia governance entities and the novel methods of disaggregated content moderation in Reddit illustrate a real desire to get ordinary users involved in some of these decisions at least some of the time. We also have seen some interest from companies in subjecting themselves to external scrutiny in a way that can be localized yet is separate from states. One example is TikTok’s creation of a collection of regional “safety councils,” which exist in the United States, Europe, Asia Pacific, and the MENA region (*Arab News* 2022).<sup>33</sup>

As for governments, well, again, we have evidence of the willingness of governments to experiment with novel forms of disaggregated authority, especially when they don’t need to give up any of their ultimate power. The European Union itself is a key example, as are the various participatory arrangements described by Fung and Wright (2003), as well as the entire apparatus of existing multistakeholder arrangements in contexts like environmental governance and even internet governance at lower network layers.<sup>34</sup> As Kaye (2019a, 146) suggests, multistakeholderism is “the overriding organizational principle for internet governance.” Indeed, some commentators have suggested that internet governance inherently takes a multistakeholder form, because of its broad range of impacts across national borders.<sup>35</sup>

While “multistakeholder governance” is often used to refer to governance arrangements that incorporate existing aggregate entities, such as companies, governments, and NGOs, there have also been some cases where so-called “crowdsourcing” – that is, the participation of ordinary people – has been incorporated. Thus far, such “crowdsourcing” has, according to at least one study, had a limited impact (Radu, Zingales, and Calandro 2015) – but to my mind, this counsels increasing our institutional ambition in order to more effectively empower ordinary people in multi-stakeholder arrangements.

Similarly, with respect to the involvement of workers, the lessons of the labor movement suggest both that workers will be willing to participate in the governance of their own workplaces when given the opportunity and that at least under some political circumstances governments can be convinced to help. The growing consciousness of workers in roles associated with platform governance of their own importance (and hence potentially readiness to participate in checking platform power) is evidenced by the growth of organizations like the Trust and Safety Professional Association and the Integrity Institute.

<sup>33</sup> Even in the absence of a will to implement proposals like those in this chapter, such regionalization could be a future path for the Meta Oversight Board as well, perhaps in the form of intermediate regional “courts” capable of introducing local interests into the adjudicative process to be taken into account at the final “appellate” level.

<sup>34</sup> Several of the chapters in Haggart et al. (2021) generally describe the debates around multistakeholderism on the internet.

<sup>35</sup> See references given by Van der Spuy (2017, 19).

Other scholars have proposed a notion of “co-regulation” according to which companies enforce their own policies, but are subject to government oversight via, for example, legislative approval of platform rules or a right of appeal to governments for oversight on a case-by-case basis; Marsden et al. (2020), for example, propose such a scheme for disinformation regulation and observe that it has been used in other contexts in European consumer protection law. Di Porto and Zuppetta (2021) have similarly proposed a co-regulation scheme to regulate platform-to-business disclosures for transactional platforms like Amazon. As early as 2009, Weiser (2009) proposed such a model for connectivity regulation. Keller (2022) reviews a number of similar proposals, while criticizing the limited vision of co-regulation, which, she (aptly) argues, fails to account for the role of actors other than companies and governments (an objection that does not apply to the broader proposals in this book).<sup>36</sup>

#### 6.5.1 *Recognizing and Mitigating the Limits of Platform Constitutionalism*

That being said, while liberal governments may tolerate or outright promote such experiments, it is doubtless true that autocratic governments have strong reasons to resist or subvert the participation of their citizens in any such system, especially as applied to social media. The last thing that a government interested in censoring its political opponents on social media and distributing its own propaganda wants is direct citizen participation in the regulation of social media content outside of its control. And while I have argued for efforts to directly recruit the participation of individuals living under the thumb of autocratic governments, this may not be good enough. Autocratic governments have immense resources to monitor and intimidate their people. Unfortunately, the benefits of institutions like those I have sketched will likely be substantially reduced when applied to situations that are in tension with the self-perceived interests of dictators: such is the world in which we live.

Another difficult problem is convincing ordinary people to participate, particularly those from relatively disadvantaged groups who may lack the time to participate, or the social capital to be effective. Existing experiments in participatory governance in political states have been successful relative in part to their capacity to facilitate the delivery of concrete, material, goods to their participants (e.g., Baiocchi 2003, 64); such benefits are obviously unlikely to be delivered in the specialized context of platform regulation. This is why I propose the relatively brute instruments of sortition to select participants (to mitigate, although not completely eliminate, self-selection by the most resourced or even corruptly motivated), along with raw financial payments, which can be calibrated over time to the point that the rate of refusal is similar among the most and the least well-resourced.

<sup>36</sup> See also Cohen (2019, 187), who criticizes co-regulation, and the “new governance” more broadly, on similar grounds.

The lack of financial resources to participate has been noted as an impediment to participation in previous multistakeholder internet processes (Belli 2015, 11–12; Hofmann 2016, 33.); moreover, brute payment has been famously effective in at least one case of thick ordinary person involvement in complex collective decision-making processes – I refer once again to classical Athens.<sup>37</sup> It may also be the case that the diminishing marginal utility of wealth may have the happy side effect of improving the inclusiveness of paid participation, insofar as persons from less developed countries and from more disadvantaged groups are likely to be more benefited by payments (and hence more likely to be motivated to participate), holding the payment amount equal.

As for procedural inequalities between groups with greater and lesser social, cultural, and educational capital, no design can perfectly remediate those differences. However, my recommendations above consciously take them into account by making deliberative training available but optional and providing that local councils have full autonomy over their own decision-making processes with the idea that they can choose culturally meaningful and familiar procedures.

A related problem is that group decision processes which are too open-ended can descend into a kind of oligarchic bureaucracy. This has been observed with Wiki contribution processes (Shaw and Hill 2014); scholars have also noted its applicability to other forms of multistakeholder governance (Hofmann 2016, 34). This seems to me to be a structural problem with participatory processes: The burdens of managing input from lots of people, in terms of the sheer number of people involved, their diversity (and hence capacity to generate novel process objections), and the increasing risk of adversarial behavior meant to sabotage rather than advance participatory processes all create pressure to manage participation via increasingly formal and rule-bound mechanisms. Such processes may exclude the most subordinated participants both because they increase the cost of participation overall (and hence exclude those least able to bear the costs at the margins), and because they are likely to privilege established and mainstream modes of advocacy (Young 2001).

An important case study of a category of risks much like the foregoing even in an elite multistakeholder entity composed of sophisticated organizational actors is Moog, Spicer, and Böhm's (2015) analysis of the Forest Stewardship Council, a transnational sustainability labeling scheme designed to represent both corporations and civil society organizations in negotiating over and shifting incentives in forest management. The FSC would appear to have a model design – one organized to overrepresent civil society and social groups relative to corporations, and in which conscious efforts to include the Global South were built into its structure from the start (Moog, Spicer, and Böhm 2015, 474, 477).

<sup>37</sup> In recent research, Fan and Zhang (2020) conducted an experiment on using juries for content moderation decisions. Their participants (recruited from Americans over Mechanical Turk) reported a substantially higher willingness to participate if paid.



Yet the cost of monitoring compliance with its labeling rules forced the organization to increasingly rely on for-profit organizations hired and paid by corporate members. In a story that will sound familiar to those who know the history of American accounting scandals in the 2008 financial crisis, those monitoring firms find themselves at the center of conflicts of interest undermining the overall compliance with certification standards. Moreover, the Forest Stewardship Council's own lack of resources has impeded it from effectively carrying out that monitoring function itself, and inequalities between corporate and nonprofit members have disabled the nonprofits from doing the same (Moog, Spicer, and Böhm 2015, 480–81). Finally, because the labeling scheme is subject to market constraints – other environmental labeling schemes exist, and so FSC labels must compete in the marketplace – nonprofit resources are further drained by the need to exert political pressure as well as deal with the demand to compromise certification standards to meet the competitive environment. Lessons from the Moog, Spicer & Böhm analysis suggest an imperative both to adequately resource participatory platform governance organizations and to insulate them from market forces, perhaps by increasing the involvement of political states and ensuring compulsory participation via law.

Ultimately, any system like that described in this chapter would have to be initiated by governments. This is true even if companies largely wanted to participate voluntarily, as governments would also have to shelter those companies from anti-trust scrutiny for unifying core governance features of their business operations in the manner just described.

Moreover, no government is likely to be unable to implement such a system singlehandedly. Platforms, especially but not exclusively social media platforms, are an international problem; moreover, in disputes between individual governments and platforms, in many cases, the platforms have sufficient negotiating power by threatening to leave or reduce services in a national market in order to procure regulatory changes, as in the dispute between Facebook and Australia over its payment-for-news legislation, in which the company forced concessions from the government by ceasing to carry news stories in the country (Meade, Taylor, and Hurst 2021).<sup>38</sup> But states have market power too, in the sense that there are certain very large markets (Europe, the United States, and India) which would be extraordinarily costly for companies to exit; moreover, smaller countries may combine to acquire market power in the same sense. In addition, the divergent national interests of different

<sup>38</sup> By referencing this example, I don't mean to suggest either that Facebook's behavior was inappropriate or that the news regulation was a good idea. In fact, it seems reasonable to me for a company to choose not to carry a particular kind of content for which a price is demanded – whether that price is from some private seller or from the government – rather than pay; I also think the Australian legislation is truly terrible policy, and that if the Australian government wants to transfer wealth from social media companies to journalists (which might be a good idea), it should do so in the ordinary fashion, that is, with taxes and transfers. But power such as that which Facebook implicitly exercised by exiting the news market in Australia can also be used under much less sympathetic circumstances.

countries are most effectively reconciled by a negotiated framework. Accordingly, in the first instance, platform governance development programs ought to be implemented by international treaty, and understood as a form of human rights treaty-making process.

For the sake of the United States, as the most significant market as well as the market with the most severe constitutional free speech restrictions, it would be advisable to frame mandatory governance institutions as wholly content neutral, and indeed even neutral between communicative platforms and transactional platforms – that is, as regulations of the abstract structure of governing user behavior on networked platforms.<sup>39</sup> Other important markets, such as the European Union, have greater capacity to regulate with fewer legal constraints, and have also demonstrated a greater willingness to regulate in their existing legislation.

In addition to establishing the framework for such a system, however, there are a number of other positive interventions that governments could take. I will conclude this volume by turning directly to states as such and considering the potential for additional governmental intervention in the platform economy in aid of more effective and legitimate governance.

<sup>39</sup> I say that the United States is the most significant market in view of the combination of its wealth, population, privileged position in infrastructural internet governance (see generally DeNardis 2009, 2014), and territorial control over the brunt of the headquarters, senior executives, and other personnel of the major existing platform companies. China could compete with the United States for that title (as with most others), but for the fact that it has excluded a number of the major platforms from operating in its mainland and runs a significant alternate ecosystem. This rules it out of consideration for present purposes. So, of course, does the fact that its government is a brutal human-rights-disregarding despotism.