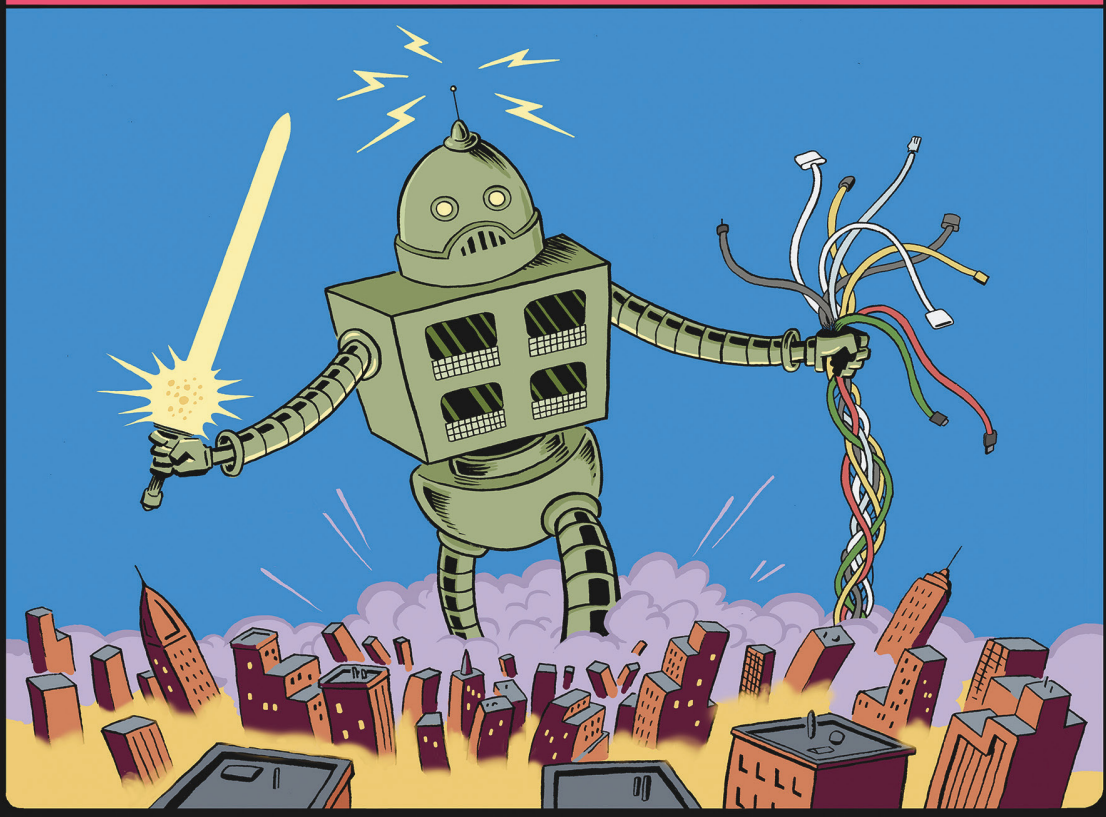
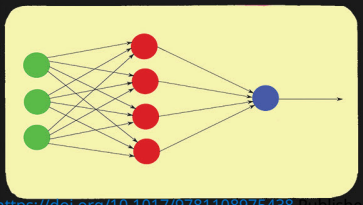
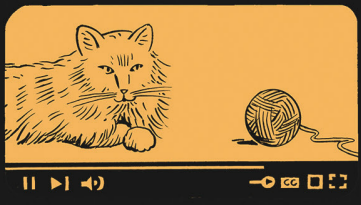
THE NETWORKED LEVIATHAN



FOR DEMOCRATIC PLATFORMS



PAUL GOWDER



https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

THE NETWORKED LEVIATHAN

Governments and consumers expect internet platform companies to regulate their users to prevent fraud, stop misinformation, and avoid violence. Yet, so far, they've failed to do so. The inability of platforms like Facebook, Google, and Amazon to govern their users has led to stolen elections, refused vaccines, counterfeit N95s in a pandemic, and even genocide. Such failures stem from these companies' inability to manage the complexity of their userbases, products, and their own incentives under the eyes of internal and external constituencies. *The Networked Leviathan* argues that countries should adapt the institutional tools developed in political science to democratize major platforms. Democratic institutions allow knowledgeable actors to freely share and apply their understanding of the problems they face while leaders can more readily recruit third parties to help manage their decision-making capacity. This book is also available Open Access on Cambridge Core. For more information, visit https://networked-leviathan.com.

Paul Gowder is a law professor and political scientist who has written extensively on the problem of holding lawful systems of governance accountable to their constituents. He is the author of *The Rule of Law in the Real World* (2016). He has also served as a consultant in Facebook's Civic Integrity and Public Policy Research teams, where he participated in the design of the Oversight Board, and is a Founding Fellow and board member of the Integrity Institute.

The Networked Leviathan

FOR DEMOCRATIC PLATFORMS

PAUL GOWDER
Northwestern University

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108838627

DOI: 10.1017/9781108975438

First published 2023

A catalogue record for this publication is available from the British Library.

A Cataloging-in-Publication data record for this book is available from the Library of Congress

ISBN 978-1-108-83862-7 Hardback
ISBN 978-1-108-97190-4 Paperback

Contents

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

Acknowledgments

This book has spanned upward of three years of writing, and my thinking about the problems of platform governance began well before this book was even conceived, in 2018, when I was embedded in Facebook's Civic Integrity team. In that time period, many people have influenced the course of this book, though, as is customary, none are responsible for its many flaws.

First, the Knight Foundation generously provided a grant to enable me to find the time to do the writing – without this assistance, the book would have been even further behind than it is. My thanks to John Sands and Anabelle Torek at Knight for their faith in the project, and to Derek Gundersen and Kevin Webb here at Northwestern for their work to handle the details of grant administration. The project was also supported by the Northwestern Pritzker School of Law Faculty Research Program. It was initiated back when I was at the University of Iowa, and I thank the Dean there, Kevin Washburn, for originally bringing the Knight Foundation's solicitation to my attention, as well as my current dean at Northwestern, Hari Osofsky, for continuing support. And – on the subject of the extraordinary amount of time that this book has taken – I thank Matt Gallaway for his seemingly endless well of patience and encouraging words as the span between the date written down in the contract and the date on which the manuscript was actually delivered grew longer, and longer, and longer still. Mandy Keifetz, who did the indexing, exercised similar levels of patience and toleration – this is the second of my books for which Matt was the editor and Mandy the indexer, and I am endlessly grateful to both. I must also express my gratitude to Wayne Shellabarger, the artist who provided the cover illustration. Wayne is not only a brilliant artist but also a seasoned professional who was an utter delight to work with.

I have also benefited greatly from the efforts of a number of student research assistants, including (in alphabetical order) Peter Chan, Chris Darby, Julia Gabriel, Janani Hariharan, Rohun Reddy, Leah Regan-Smith, and Leo Soh. The Northwestern Pritzker School of Law has the best librarians in the business, and Sarah Reis and Tom Gaylord provided irreplaceable support in getting this work achieved, as did Sarah Shoemaker and Jane Brock from Northwestern's faculty support staff.

This book has also benefited from a variety of presentations, workshops, and meetings at which part or all of its contents were discussed, including a workshop at the University at Buffalo School of Law's Baldy Center, the 2021 conference of the Tilburg Institute for Law, Technology, and Society (TILTing Perspectives: Regulating in Times of Crisis), the 2021 First Annual Conference of the Platform Governance Research Network, Albert Yoon's seminar, "The Good, Bad, and Ugly of Academic Research" at the University of Toronto in Fall 2022, and various internal conversations at Northwestern, including a discussion with the J.D./Ph.D. seminar and an early presentation of the basic idea. I thank the participants in all of those events. Most significantly, Northwestern's Center on Law, Business, and Economics organized a one-day manuscript roundtable for the penultimate draft of this book which provided critical feedback – I thank Matt Spitzer for inviting me to participate in this process, and Derek Gundersen (once again) for implementing it.

A number of individuals, either through those workshops or on their own, have offered particularly helpful feedback and advice, including Karen J. Alter, Chinmayi Arun, Julie Cohen, Zoe Darmé, Jennifer Forestal, Seth Frey, Gillian Hadfield, Katie Harbath, Aziz Huq, Andy Koppelman, Sahar Massachi, Naomi Shiffman, Jim Speta, Vanisha Sukdeo, Heidi Tworek, and Ari Waldman.

I would also be remiss not to thank the members of the Integrity Institute as well as the former Facebook Civic Integrity and Product Policy Research teams for many years of conversations about the problems of platform governance. I particularly appreciate Nicole Bonoff, who first brought me into the room by recruiting me to work with Civic Integrity.

Leonidas, of course, kept me (relatively) sane through the process, and Drew kindly and supportively tolerated the repeated bouts of derangement it occasioned.

Introduction

The Perils of Platform Misgovernance

There's no way that Mark Zuckerberg could have imagined, when he created an electronic version of Harvard's traditional collection of student photos – informally called "the facebook" at least since I was in law school there in the 1990s – that a decade and a half later he would have the blood of genocide victims on his hands.

In 2016 and 2017, the Myanmar government carried out a genocide against the country's Rohingya Muslim minority.[1] At the time, Facebook was the overwhelmingly dominant internet platform in Myanmar, with one report commissioned by Facebook itself observing that "[t]here are equal numbers of internet users and Facebook users in Myanmar" and "many people use Facebook as their main source of information."[2] This is in part because Facebook and local providers created subsidized forms of internet access such as "Free Basics" and "Facebook Flex" in order to expand their services in the country.[3] Instrumental in the atrocities was propaganda which the military distributed over Facebook.[4] Facebook acknowledged its role in 2018.[5]

The shock and chagrin that Zuckerberg must have felt when he learned of his culpability in the ethnic cleansing of the Rohingya people was probably pretty similar to how the leaders and founders of every major social media company

[1] Domino (2020, 150–1); United Nations Human Rights Council (2018); Beaubien (2018); *BBC News* (2020). On Facebook's culpability, see the independent assessment Facebook commissioned in 2018 of its human rights impact in Myanmar, by Business for Social Responsibility, published at https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf in October 2018. The United States Department of State formally and publicly classified the events in question as a genocide in March 2022, only the eighth time since the Holocaust that it has made such a declaration. US Department of State, Genocide, Crimes Against Humanity and Ethnic Cleansing of Rohingya in Burma, www.state.gov/burma-genocide/ (last visited December 4, 2022).

[2] Warofka (2018, 13).

[3] United Nations Human Rights Council (2018, 339–40); for further context, see Stecklow (2018).

[4] Amnesty International (2022); *BBC News* (2018a); United Nations Human Rights Council (2018, 340–1); Hogan and Safi (2018); Gowen and Bearak (2017); Mozur (2018); Fink (2018). Other media were also implicated, such as the official state media of the country, Lee (2019).

[5] *BBC News* (2018b).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

felt on January 6, 2021. After years of the weaponization of misinformation and polarization by Donald Trump and his allies over social media, his regime finally culminated in an armed mob attack on the US Capitol for the purpose of preventing the peaceful transfer of power. While the major social media companies have knowingly or even intentionally inflicted many terrible things on the world, it strains credulity to suggest that they intended or even expected a coup attempt in the world's richest and most powerful democracy – and not incidentally the country in which all of those companies are headquartered, from which their founders mostly originate, and under whose political order those companies and their leaders have prospered.

This book begins with those two examples, among the many ways in which social media companies have failed their users and the world, because their very extremity highlights a point that will run through this book: *nobody* (except their perpetrators) *wanted those two things to happen*. Mark Zuckerberg might be a bad guy. He may be the face of a business model, which often goes by the name "surveillance capitalism" (e.g., Zuboff 2019) and which causes numerous individual and social harms. But he's not Hitler. Imagine that a genie had appeared before Mark Zuckerberg and said "you can give up some moderate proportion, say 15%, of your wealth from Facebook and, in exchange, I will magically cancel the genocide that you'd otherwise be culpable for." I imagine that he would have taken the deal.

This isn't an apology for Mark Zuckerberg. I don't really care what you think about him. But this book does assume that he doesn't want to be responsible for genocide, and that he would be willing to spend a substantial (though perhaps not infinite) amount of Facebook's money to avoid that fate, if only he knew how to do so. The same goes for the then-leaders of all of the major companies and the coup attempt on January 6.[6] To avoid being culpable in the next genocide or coup, they need to do a better job at getting control of what happens over the services they run, that is, "platform governance."

This book focuses on problems, like genocide and coup attempts, where the interests of companies and their leaders are aligned with the interests of the rest of us – ordinary people across the world and our (decent, liberal-democratic) governments. It rests on the assumption that there is a substantial amount of social harm caused in that territory of interest alignment, such that companies have reasons (whether moral or financial) to work together with people and governments to redesign the services those companies provide in order to address those harms. This book sketches out one way in which we might do so.

I take no position on any of the broader questions surrounding these services and the extent to which their interests might conflict with those of the rest of us. Social media in particular has been subject to a sustained critique rooted in its

[6] Alas, I can't confidently say the same about the leaders of rogue minor companies like Parler and Gab, or about the leader of Twitter at the time of this writing, Elon Musk.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

revenue model and the way that revenue model encourages companies both to radically undermine individuals' autonomy over the details of their lives and to promote thoughtless, emotion-driven, content in pursuit of the goal of "engagement" (and thereby advertising dollars). I don't purport to evaluate that critique, or to attempt to balance the harms social media generates against the benefits of interpersonal connection and communication which it offers. First, let's solve the genocides and coup attempts, and then we can worry about surveillance capitalism. If I can contribute, even in a small way, to reducing the risk of another Myanmar genocide or another January 6, then I will consider this book a success.

In the pages that follow, I offer a program, built on the insights of political science and allied fields, for radically democratizing services like social media – more broadly, "platforms." We should directly insert ordinary people, including ordinary people from the Global South, from minoritized and indigenous communities in the North, and from other subordinated and excluded groups, directly into the processes of platform rule enforcement, rule development, and ultimately product design. I argue that such innovations would actually be in the long-run interests of platform companies along with the rest of us.

In a nutshell, here's what I will propose (with the details reserved for Chapter 6). Companies, assisted where possible and coerced where necessary by liberal-democratic states, will create a multilevel participatory governance organization, organized along lines of geography as well as identitarian affinity.[7] Through that organization, randomly selected and well-paid groups of ordinary people will have (a) access to internal company information; (b) privileged channels of communication to companies about their own observations of the local impact of platform policies; (c) some degree of retail-level control over company governance decisions with respect to users (e.g., in the social media context, appellate authority over content moderation decisions); and (d) some degree of wholesale-level control over company policies through the ability to propose (and in some cases veto) rule and product design changes.

This proposal should be backstopped by certain legal interventions, including the judicious use of antitrust and workplace rights law and the subjection of the largest companies to elements of the international human rights regime (mainly the responsibility to protect). However, at its heart, the incentive for companies to participate is that doing so is in their own interests. The tasks that fall under the rubric of "platform governance" as defined in this book – at least within the domain in which company interests and public interests are aligned – are tasks that companies actually need to carry out well in order to protect their own long-term economic viability. But, this book argues, they are persistently hampered in doing so

[7] That is, I envision local and identity-based first-level groups – also including company employees as individuals – which nominate members to composite second-level groups with more authority, and so forth. Company participation in this system will begin with larger companies and social media companies, with expansion to smaller companies and other types of platforms over time.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

by (a) their inability to give people who are socially distant from their personnel the capacity and incentive to supply them with the information they need to frame and implement their rules, and (b) their self-control problems in resisting short-sighted incentives relating to threats (especially from governments) and destructive methods of short-term profit seeking. Giving up some control to ordinary people across the boundaries of geography, hierarchy, and affinity can alleviate both problems at once: their self-control problems by separating those who care about short-term costs and threats from those who control the decisions which implicate them; their information problems by creating channels of communication between the peripheries of a company's domain and its organizational nerve centers as well as making the use of those channels effective at achieving the ends of those at the peripheries, and hence worth doing for them.

A major challenge in writing a book-length work on academic timescales in this domain is that the platform economy and its surrounding social, political, and legal landscape tend to change at astonishing speed. Between the penultimate draft of this manuscript and its final, for example, Elon Musk acquired Twitter, at which point the company went from having some of the most thoughtful and innovative work on platform governance to operating purely arbitrarily.[8] Alternative platforms, some with radically different organizational structures (such as the Mastodon federated model), suddenly became popular. Around the same time period, the Fifth Circuit issued a decision purporting to uphold a Texas law prohibiting social media platforms from engaging in "censorship" on the basis of viewpoint – obviously motivated by efforts by politicians who are supported by the extreme right to require platforms to host misinformation and hate speech.[9] By the time this book sees print, the Supreme Court may have reversed that decision – or it might be upheld, and the enterprise of social media content moderation may be effectively dead, at least as it relates to partisan misinformation. That sort of radical change seems to happen every time one looks up from one's keyboard when one tries to write about platforms.

This book aims to be timely insofar as it takes account, as best as possible, of the current state of play in the operation of platforms and offers advice that is potentially actionable by existing governments and companies for mitigating their governance problems. However, it also aims to be timeless – and hence unlikely to become obsolete the moment it hits the shelves – insofar as it offers a general account of the sorts of governance problems that platforms face, regardless of which companies

[8] Incidentally, unless otherwise specified, discussions of specific governance features or policies of Twitter in this book refer to the state of affairs before the Musk acquisition, when the company was making a serious effort to conduct platform governance. At the time of this book's completion, matters on Twitter under Musk are too chaotic to fully take into account.

[9] *NetChoice v. Paxton* No. 21-51178 (5th Cir., September 16, 2022). That decision was, to put it bluntly, utterly clueless – its analysis of companies' First Amendment interests completely neglected the well-recognized role of content moderation in their core business models.

happen to be in operation on a specific date, and an account of how that arises from the nature of the services they provide. From that perspective, the descriptions of individual companies and incidents given in this book should be taken primarily as evidence for the general structural account of the problem. I hope that in the future this book may even serve as a contribution to the discipline of political science more broadly by providing an illustration of the incidence of state-like problems of governance in nonstate contexts.

The price for this effort to speak to the present as well as the future and to policymakers and company personnel as well as to scholars is that none of its audiences will be fully satisfied. Policymakers and company personnel will have cause for complaint that this book lingers far too long over excursions into the scholarly literature, and its recommendations will not be finely tuned to their most immediate problems but will require adaptation to be useful. Scholars will have cause for complaint that this book sacrifices the extended theoretical development of fewer ideas and in-depth interaction with the literature in favor of a somewhat more brisk development of more ideas that can come together into relatively concrete recommendations to capture that low-hanging governance fruit which I just described. To each of those groups, I acknowledge your complaints and trust and hope that the payoff – real design and policy recommendations backed by decades of scholarship for officials in companies and states, and a novel application of existing tools in political science to new contexts for scholars – warrants your patience.

WHAT IS "PLATFORM GOVERNANCE" ANYWAY?

Before we get going, I should specify the scope of the argument. This is a book about platforms and about governance, and the merger of those two things that has come to be called "platform governance." Standing on its own, however, that sentence indicates surprisingly little because of the capaciousness of all those ideas.

First, "platform." I'm going to hazard a more careful definition of a "platform" in Chapter 1, but for present purposes, we can take it to mean internet enterprises that facilitate interactions between individuals, where the economic model of the enterprise is that the company operating the platform captures some of the value produced by that interaction.[10] This description captures two-sided markets like eBay, Uber, Apple's App Store, and Amazon's Marketplace, which take commissions on transactions; as well as social media companies like Facebook and Twitter, which

[10] This is generally similar to other extant definitions of platforms in the scholarly literature across several fields. Bonina et al. (2021, 871) helpfully review recent definitions along these lines, as does Jin (2017, 7–10). In the terms of Bonina et al., this book focuses on "transaction" platforms (which encompasses both social media and many-to-many marketplaces like Amazon and eBay) rather than "innovation" platforms, although I shall reserve the term "transactional" for platforms that primarily focus on buying and selling rather than social interaction. There has also long been talk of hardware "platforms," such as the iPhone. Those are entirely out of scope for this book.

capture some of the value from users producing and viewing one another's content by sticking ads on the top and mining the data they generate. In this book, I'm primarily concerned with the social media kind, with which I'm most familiar and which seem to lately be causing the direst social impacts. But many of the propositions advanced in these pages will be oriented toward the abstract characteristics of platforms and will apply reasonably well to the transactional kind too.[11]

The difference between transactional and social media platforms has the potential to vex any academic work attempting to analyze them together. One challenge which might fairly be raised against the entire project is to question whether they even bear enough in common to be understood together – do commentators and scholars just lazily use the word "platform" to describe both kinds of enterprises? Are not the problems posed by each – such as evasion of ordinary commercial regulation for the transactional kind and political polarization and disruption for the social kind – distinct?[12]

I will attempt to answer that worry through the back door with the additional definitional work in Chapter 1, but here we might just notice that many of our major companies have demonstrated a striking tendency to leverage their assets to operate across these business models. For example, Meta's major properties, Facebook and Instagram, both operate transactional marketplaces – Facebook's focused on eBay-like individual-to-individual transactions, and Instagram's on business-to-consumer transactions.[13] I submit that the reason that this seems like a good idea is because the core properties of a successful social media company, in terms of the scale of its userbase and the incentives it offers for activity, are also the core properties of a successful transactional platform. Moreover, the basic criteria for successful design are similar across these platforms. For example, Gillespie (2018b, 211–12) identifies that these two types of platforms offer something like algorithmically curated and moderated user-generated content within the core of their business.

As a first pass, we can say that "platform" is an abstract description of an economic model focused on facilitating a variety of kinds of third-party interactions, while "social

[11] The big Chinese platforms, such as Sina Weibo and WeChat, are outside the scope of this book. The goals and challenges of and tools available to a largely single-state platform operating under the thumb of a world-historically sophisticated and effective autocracy are wholly distinct from those of a platform with a global userbase operating out of a liberal democracy.

[12] For example, Lobel (2016, 94–95) suggests that social media is merely a "prelude" to the real platform companies, like Uber.

[13] Incidentally, it is quite inconvenient that two of the major platform companies, Facebook and Google, changed their names while maintaining the original name for a subset of their original businesses, during the course of the events described in this volume. Generally, I will use "Facebook" and "Meta" interchangeably, and the same goes for "Google" and "Alphabet." However, I will endeavor (with only moderate consistency) to primarily use "Meta" for references to elements of the company formally known as Facebook in contexts of continuing operation, for example, the "Meta Oversight Board" rather than the "Facebook Oversight Board." It should be clear from context when I mean to refer to Facebook the service in contrast to other Meta services like Instagram, and to Google the service in contrast to other Alphabet services like YouTube.

media" describes a particular kind of interaction facilitated by some platforms. To the extent the problems of governance in social media are rooted in the abstract economic model, they will be shared with other kinds of platforms like Amazon and eBay; however, some problems of social media governance arise in particular from the *communicative* character of the interactions social media platforms facilitate.

I also think it is a mistake to be too strict about one's definition of "platform." In reality, "platform" is probably a Wittgensteinian family resemblance concept rather than susceptible to definitive criteria of inclusion and exclusion.[14] There are characteristic properties of platforms, but not everything which we might want to call a platform will have all those characteristic properties. For example, the outer limits of the concept of a social media platform on the definition I favor are probably at Google Search, which I consider to fit within the definition insofar as users are both creators of and consumers of search results. (They are creators both by creating the underlying web pages and other content and by creating the links and other activity which Google relies upon for search rankings.) Moreover, Google Search experiences the characteristic problems associated with social media, such as user gaming of recommendation algorithms to promote low-quality content. Finally, Google Search monetizes activity in exactly the same way that Facebook or YouTube does, that is, by using behavioral data to create recommendation algorithms that give users an incentive to make more use of their service, and by using predictions from that behavioral data to target advertisements.[15] In order to distinguish platforms from more longstanding business models such as brokers of various sorts, it will also help to focus our attention on the kind of novelty that platforms generate – platforms tend to create new kinds of interactions between people or transform existing kinds of interactions in fairly dramatic ways.

Why and How Do Platforms Govern?

This leads to the notion of "governance." There are at least two distinct problems of "platform governance," though, as I shall argue, they are closely related. First is the governance *of* platforms, that is, of platform companies, by governments.

[14] I mean to invoke the weaker version of the notion of a family concept as described by Wennerberg (1967, 109–10).

[15] However, some important clusters of network affordances, some of which are even controlled by platform companies, probably don't meet any formal definition of "platform" which we might want to adopt. We might call them "quasi-platforms." For example, WhatsApp seems to have many of the characteristic problems of social media platforms, such as viral misinformation, but it lacks many of the standard features – because data flowing over it are encrypted, Meta has limited opportunities at best to monetize those data; as far as I know, it doesn't feature recommendation algorithms in any significant sense; the fact that it does seem susceptible to things like viral misinformation seems, as far as I can tell (and this is with extremely low confidence), to be a product of the combination of its dominant position in certain communicative markets as well as more user-interface style affordances like the capacity to forward messages to many people at once and, perhaps, a certain degree of immersiveness not shared by, for example, email. At any rate, we can pretend that WhatsApp is a platform to the extent the governance techniques described in this book might be useful for it, but ignore it otherwise.

There are numerous debates about the extent to which we should regulate those companies, and the manner in which we should do so; for example, how much should we require them to protect the data of their users, or forbid them from combining that data in certain ways? Should they be required to offer users some degree of interoperability in terms of being able to move data back and forth between them?[16]

However, there is also the problem of governance *by* platforms – that is, the regulation of the behavior of users of platforms, such as buyers and sellers in Amazon marketplace or people posting and reading tweets.[17] It is now commonplace to recognize that platforms are engaged in acts analogous to public governance when they regulate user behavior – thus, for example, many scholars have identified the quasi-governmental character of social media speech regulation, and have further identified that this is intrinsic to the products companies offer (Klonick 2018, 1638–30; Gillespie 2018b). One scholar (Eichensehr 2019) has gone even further, comparing companies to "Digital Switzerlands" that compete for authority with physical-world governments; a claim that may seem overheated but for the fact that the author leads her article with a citation to the president of Microsoft arguing for just such a thing (albeit in the limited domain of protecting their customers from cybersecurity threats).[18]

In the process of exercising governing power, platform companies routinely adopt both the methods and the personnel of government regulators, hiring lawyers – many with prosecutorial or other government experience – to write lawlike rules which they enforce with formal processes. The apotheosis of this trend is perhaps Meta's Oversight Board, colloquially called its "Supreme Court" since before it was even created.

This degree of government-like organization and government-like behavior is an outlier in contemporary capitalism. The point should not be overstated: As Rory Van Loo (2016) has shown, corporations frequently offer dispute resolution services (consider credit card chargebacks as the canonical example). But the degree of lawlike formalization in the platform economy seems unique.[19] When my bank decides whether or not to offer me a line of credit, I don't have the benefit of a

[16] I use "governance" and "regulation" interchangeably.

[17] See Gorwa (2019a, 855); Gillespie (2018c) for the governance of/governance by distinction.

[18] See also Srivastava (2021, 7–8), who articulates a similar idea from an international relations perspective; and Cohen (2019, 129–31) giving an example of a tug of war over surveillance which positions companies as both defenders of the public against state surveillance and themselves agents of both private and state surveillance.

[19] Another important potential counterexample is supply chain regulation, in which companies control the behavior of their suppliers for social goals (albeit driven by their business interests, such as consumer demand), in fields such as labor rights and environmental protection. Green (2014, 1–2) cites Walmart's sustainability rules as a major source of regulation for numerous global producers. Van Loo (2020) gives a variety of other examples in that vein. Still, regulating the other business firms with whom one deals is a different ballgame from regulating a mass public.

published set of rules to which I may appeal in litigating their decision before a formal process within the company. Shopping malls don't have their own codes of laws to justify the decisions of their security personnel to kick out rowdy shoppers. Even credit card chargebacks don't feature anything like a system of appeals or a published set of rules meant to draw a balance between the interests of customers in getting what they purchased and the interest of retailers in avoiding fraud and manipulation.

Notably, these other businesses offer their customers less of the protection of formal legalism even though the individual stakes for those customers are typically much higher than at least the social platforms. If I have one of my Facebook posts removed, I typically don't experience a serious injury.[20] By contrast, if some merchant defrauds me and Chase refuses to reverse the charge on my credit card, I could lose lots of money. This suggests that the increasing formality of platform user-governance decisions is not simply a response to the stakes involved for their users.

Rather, I claim that lawlike, formal, governance methods respond to the sorts of decisions that platform companies have to make. Lawlike forms of governance aren't just chosen at random; rather, human societies have developed social technologies in the most foundational sense, like the independent[-ish] judge and the written code of laws, because those are effective at solving certain kinds of governance problems (Gowder 2016, 40, 59–62; 2018b, 91).

The existence of platform law and platform law enforcement is a kind of return to an (alleged, albeit highly controversial) earlier day of weak states in which private law was required to fill in the gaps for the purposes of facilitating things like trade.[21] With that precedent in mind, it's easy to start by observing that part of the explanation for the lawlike form of platform governance proceeds from the global nature of the problems posed by the largest of such platforms and the difficulty domestic governments have in controlling them (consider our experience in the United States with counterfeit products from China and election interference from Russia).

But global scale isn't the only reason these platforms occupy a kind of governance role, and there are lots of global companies that don't. Rather, I think it is the combination of scale, vast diversity (explored further in Chapter 1), and the fact that platforms inherently (indeed, as part of the definition I've lightly sketched so far) create surfaces for interactions between third parties which by their very nature

[20] While I might experience a serious injury if I ask a company to remove someone else's post and the company says no (e.g., if that other person's post includes my private information), most of the lawlike protections social media companies offer are directed at protecting the interests of the poster, not the complainer. For example, when the Meta Oversight Board was first created, there was no way to appeal the company's refusal to take down content someone else produced, only the company's decision to take down content one has produced. (This appears, however, to have changed.) Similarly, there's an appeals process for YouTube creators to seek review of platform "strikes," but as far as I can discern there is not one for people who report policy-violating videos.

[21] For example, Milgrom, North, and Weingast (1990); for a skeptical take, see Kadens (2015).

enable and promote some interactions and disable and deter others that drives the phenomenon. The need to regulate (i.e., govern) the behavior of the parties using their platforms, at least in their own interests, is built into the economic model itself: For platform companies to make money, there must be activity to monetize; for that activity to be sustainable in the long term, users must on the whole understand themselves to be experiencing positive outcomes from their usage of a platform.[22] But human sociality has the unfortunate habit of turning vicious on a regular basis, and a platform too-plagued by viciousness (as its users understand it) will not be able to keep its users. For the simplest example: if Amazon lets buyers and sellers rip one another off with abandon, before long there won't be any transactions on their platform for the company to take a cut of.

That is the core dynamic driving the distinctive enterprise of "platform governance" and it's the reason that things like credit card chargebacks don't count: Amex doesn't have to worry that people will stop using its credit cards if merchants rip them off, because the credit cards don't (except in the special case of things like skimmers) have any particular connection to the rip-offs – a person sold a piece of garbage by their local store would have gotten just as ripped off if they'd paid in cash. Accordingly, Amex's business model doesn't directly depend on governing transactional honesty (though it might be able to obtain a competitive advantage by doing so, and doing so effectively, or government regulators might impose it) in the same way that Facebook's and eBay's business models depend on governing what shows up on their platforms.

"Platform governance" in the sense of governance-by-platforms is helpfully divided into three subcategories, which we can, at a first pass, call "organizational governance," "architectural governance," and "regulatory governance."

In the first category falls choices made about the internal structure and processes of an organization itself, such as the organization of decision-making responsibility among employees for the exercise of authority over behavior using the platform. Traditional "corporate governance" falls within this category, but so do novel innovations such as Facebook's creation of a content moderation oversight board (Douek 2019; Klonick 2019).

In the second category would be what Lessig (1999) described as the regulatory capacity of code or architecture. While code may (and often does) unintentionally regulate, we should limit terms like "architectural governance" to the intentional modification of the affordances made by networked internet platforms in order to control behavior. For example, Facebook has experimented with restricting the ability of non-posters to see the "like" counts on posts, apparently in order to

[22] I leave aside here the problem of platform addiction, which could be understood as users fearing negative experiences from leaving the platforms to which they've been habituated. However, the possibility of addiction or of less psychological and more economic analogues (such as lock-in due to high switching costs for platforms that provide important services) imposes some limit on the scope of this claim.

reduce the behavioral incentives supplied by the visibility of "likes" (Constine 2019). One of the persistent features of platforms due to their artifactual character is the fairly blurry boundary between product design and governance (further discussed below).

Finally, the third category describes the most explicit forms of platform governance, by which platform operators make (stated or unstated) rules that divide conduct into permissible and impermissible behavior. Such rules may be enforced by human beings making decisions or by automated processes, and can be backed by sanctions, such as the removal of offending content, offending products or listings on transactional platforms, or offending users altogether ("bans") as well as a variety of other "remedies," such as the downranking of content in algorithmic feeds or the removal of a user's capacity to earn money on their content (Goldman 2021).

Of course, the boundaries between these categories are inevitably fuzzy and unstable. Sometimes internal governance structures may be compelled by external actors, or may be adopted in order to stave off regulation by those actors. Facebook's content moderation oversight board could be described as a form of organizational governance or a form of regulatory governance. Even architectural governance is not completely bounded, as, for example, the affordances available to users may be modified in the course of a platform's exercise of regulatory governance, as when a platform disables certain features in order to control user behavior; moreover, architectural governance is backstopped by law itself.[23] Nonetheless, keeping the three broad categories of platform governance in mind will help in clarifying our thoughts about the options for platform operators and states.

Another area of unavoidable ambiguity is the boundary between governance and ordinary operation of a platform. Platform operators may make the same categories of choices in order to control user behavior that is perceived to be harmful and in order to optimize for other desirable qualities. For example, Google may choose to rearrange the ranking of websites in its search results, or Facebook of posts in its News Feed, in order to display results beneficial to their revenue models for non-regulatory reasons – that is, to display more relevant search results or more engaging posts in order to drive more usage (and hence advertising revenue). But they may also reorganize their rankings in order to prioritize behavior considered harmless over behavior considered harmful, like the distribution of viral hoax content (e.g., Constine 2018; Hearn 2017). Moreover, those motivations may merge: It might be the case that some harmful content is also detrimental to engagement or to the relevance of search results. However, they may also conflict,

[23] Law forbids the use of technical means to evade architectural restraints, for example, as in the American Computer Fraud and Abuse Act, 18 U.S.C. 1030, which forbids at a minimum traditional "hacking"-type evasion of platform architecture, and potentially could be used (abused) to forbid a much broader class of activity, such as using adversarial machine learning examples to fool artificial intelligence systems (Calo et al. 2018).

as if it turns out that the viral hoax content is also the most engaging, and, hence, the most profitable. Governance is at least partly a matter of product design, and different governance decisions can vary depending on whether those making the decisions take, for example, a long-term or a short-term perspective on the health of a product.[24]

Political Governments Can Help Platforms Govern

The point of intersection between the problem of governance of platforms and the problem of governance by platforms is that, as I shall argue in somewhat more detail in Chapter 2, one helpful way that governments and the democratic (hopefully) peoples behind them might govern platforms is by intervening on how they govern their users. Existing efforts to regulate users by regulating the platforms over which they interact are already familiar, particularly in the "intermediary liability" context most famous in the form of the US Digital Millennium Copyright Act's notice and takedown process, which does not merely dictate that companies are not to host copyrighted content but also imposes specific processes that they must offer to their users for raising and disputing claims of copyright infringement in order to benefit from a "safe harbor" provision against the companies themselves being held liable for such infringement. In effect, then, internet companies have been pressed into service (on pain of their own copyright liability) as enforcers of copyright law against their users.[25]

The example of the Digital Millennium Copyright Act also indicates some of the dangers of the government recruiting companies into a governance role. It's fairly clear that the DMCA has led to overenforcement of copyright online, at least in some respects – relying on platform enforcement is a cheap and easy method for copyright holders to get relief – so cheap and easy that it tends to be overused, and individuals with legitimate claims of difficult to adjudicate rights like fair use find themselves struggling to protect those rights. Making matters worse, some platforms have voluntarily gone well beyond even the DMCA process – for example, YouTube has a "content ID" system that proactively identifies allegedly infringing content.[26] Thus, we have recently seen the atrocious spectacle of abusive police officers playing Taylor Swift songs in order to prevent citizens from exposing their

[24] This too is not dissimilar to the governance challenges faced by states; one way to read Mancur Olson's (1993) famous article about "stationary bandits" is as an account of the governance consequences of lengthening time horizons (much more on this in Chapter 4).

[25] Generally, scholars have identified that "cooperative" governance across companies, states, and the public at large is necessary for the kinds of cross-national and complex entities that platforms are (e.g., Helberger, Pierson, and Poell 2018), while "collaborative" governance in which some discretion in governing decisions is shared between the government and private entities is common domestically as well (Donahue and Zeckhauser 2011).

[26] For a general summary of the role of platform companies in policing intellectual property on their own initiative, see Cohen (2019, 123–25).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

official misconduct – on the evident theory that if concerned citizens post videos of police misconduct on YouTube, the content ID system will take them down because of the copyrighted music playing in the background (Schiffer 2021; Cole 2021). If the abuse of platform copyright enforcement to facilitate abusive policing is too depressing, here's a more (bleakly) amusing example: In 2009, Amazon infamously discovered a rights glitch and memory-holed George Orwell's 1984 from users' Kindle devices (Stone 2009).

The DMCA can perhaps stand as the nadir of government efforts to recruit private companies to govern their users. Because US copyright law is notoriously captured by media companies (Lessig 2003), it shouldn't surprise us that an intermediary liability framework built from it would give platforms an incentive to err on the side of total overenforcement. But it could easily get worse. For example, consider the contested relationship between the notion of "terrorism" and political dissent, and the fact that US law prohibiting "material support" for terrorism is infamously overbroad, already forbidding, for example, human rights organizations from advising allegedly "terrorist" organizations even on *lawful and nonviolent* ways to achieve their political goals.[27] It seems like it's only a matter of time before we see an intermediary liability framework forcing platforms to deny services to allegedly terrorist groups or their supporters. Some scholars have argued that existing material support statutes could be interpreted to subject platform companies to criminal liability for hosting the content of such groups (e.g., VanLandingham 2017).

Similarly, a number of countries have followed a model pioneered by Germany's NetzDG law and imposed DMCA-style requirements on companies to more broadly address prohibited content.[28] There is evidence that some governments also engage in informal DMCA-like use of platforms as a kind of cat's paw to demand the suppression of content they believe to be illegal without the ordinary process imposed on states (e.g., Elkin-Koren 2022). This is similar to a variety of other ways in which governments can use private actors' control over important social affordances as a method to informally impose sanctions on individuals without complying with their own internal rules; a prominent non-internet example would be the use by some American municipalities of "nuisance property" laws to bully landlords into evicting tenants whom authorizes deem to be troublesome (Gowder 2021, 176–77). Moreover, as Citron (2018) argues, this kind of cat's-paw regulation can expand the scope of government authority not only by freeing it from procedural constraints, but also, in the social media context, by freeing it from geographic constraints as well as substantive legal constraints – she describes

[27] And somehow this is considered constitutional. *Holder v. Humanitarian Law Project*, 561 U.S. 1 (2010).

[28] See Article 19, "Germany: The Act to Improve Enforcement of the Law in Social Networks," August 2017, www.article19.org/wp-content/uploads/2017/12/170901-Legal-Analysis-German-NetzDG-Act.pdf; Zurth (2021) and commentary from the EFF (Rodriguez 2021).

how the European Union, by threatening companies with regulation, has begun to build the ability to create extraterritorial effect not only for its speech laws but even for its extralegal speech policies.[29]

The DMCA model of intermediary liability – in which governments decree what the rules are to be, and then make companies enforce them on pain of punishment – is not the way to bring together government and platform regulation. Rather, I shall argue that governments might helpfully intervene in platform governance of their users by assisting and giving companies incentives to develop robust, quasi-democratic, governance *institutions* to help create and enforce platform rules. This too is a familiar strategy for governments whose citizens are disadvantaged by the governance failures of others, as exemplified by multinational and international efforts to promote the rule of law in war-torn, transitional, developing, and failed countries. In its best form (which is sadly rarely achieved in the actual world, cf. Gowder 2016, 168–76), such projects represent efforts to actually achieve the benefits of effective government elsewhere without the cost to democratic legitimacy of simply imposing such a government on one's own. And if done right, institutions that provide for at least semi-democratic kinds of governance can also protect against the kinds of capture that the DMCA exemplifies.

Platforms Need the Help: They Are Often Unable to Govern Their Users

Platforms have shown themselves sometimes unable and (at least in part) sometimes unwilling to adequately govern the conduct of their "citizens" – that is, their users. I'll give a couple of examples from Facebook, just because it's the company I know the best, having done some work for them. (If this troubles you, please refer to the Appendix to this Introduction for a discussion of potential conflicts of interest raised by my past work with Facebook, and why you should still believe what I say.)

First unable: return again to the genocide in Myanmar. I discuss the problems that contributed to this disaster at length in Chapter 3, but in summary, we can fairly say that Facebook exacerbated the violence by its inaction, and that this inaction was attributable to the challenges associated with global scale, and the company's failure – until the crisis revealed its neglect to itself and the world – to build the capacity to engage in content moderation in the language in which the genocidal incitement was conducted. In 2021, Mozilla released a report suggesting that these problems continue on other social media platforms: Content identified by users as problematic was apparently something like 60 percent more likely "in countries that do not have English as a primary language" (Mozilla Foundation 2021). In addition to Myanmar, Facebook has struggled to prevent demagogues from using its platform to abuse their citizens in other international contexts;

[29] Balkin (2018) has described this as "new school speech regulation." See also Bloch-Wehba (2019) along similar lines.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

another prominent example comes from Duterte's abuse of the platform in the Philippines (Alba 2018; Etter 2017).

As for unwilling: There are credible allegations that former Republican operative and current Facebook Vice-President Joel Kaplan successfully blocked measures within Facebook that would have at least attempted to tackle its contribution to extremism, polarization, and misinformation in the United States, allegedly on the ground that such measures were biased against conservatives (but, alas, conservatives were responsible for more of the pernicious content).[30]

Other companies, including non-social-media platform companies, have similar problems. Amazon, for example, has notorious problems with counterfeit goods, to the point that it was moved in 2019 to list counterfeiting as an investment risk factor in its annual report (Kim 2019).

The foregoing examples also illustrate that unable and unwilling are (surprising nobody) hard to distinguish: When a platform could solve a governance problem by spending a lot of money, do we say that it's unable to do so if the solution is particularly expensive? When it bows to political pressure, do we count that as in some sense volitional? The answers to such questions can only depend on the goals motivating their asking. There is no doubt some amount of money that Amazon could spend to effectively police counterfeits, especially when sellers use Amazon's logistics services such that the counterfeit goods pass through its own warehouses. But it might be extremely costly. For example, with a truly staggering amount of money, Amazon might have experts manually inspecting every good passing through its warehouses – but I don't think it would be reasonable to describe Amazon's leaders as "unwilling" to prevent counterfeit products if the only way to do so is to spend a company-ruining amount of money.

Similarly, it took Facebook years longer to invest in tools like automated hate speech classifiers in Hindi and Bengali than it did in English (Zakrzewski et al. 2021) – had it thrown money at the problem earlier, it probably could have had those tools earlier, and may have mitigated its numerous problems in preventing violent and demagogic content in India (e.g., Frenkel and Alba 2021). But companies don't have unlimited money, and not even pre-recession Facebook could throw money at machine learning in every language on Earth. Do we call Facebook unwilling to invest earlier because it failed to prioritize languages with such huge populations on its platforms? Do we call it unable because (ex hypothesi) it would have done so with unlimited resources? Does our intuition about this change depending on whether anyone in Menlo Park knew of the dangers posed by hate speech in India? Whether anyone in Menlo Park *should have* known about those dangers, for example, because they should have had better ways to learn about them, again, given the gigantic number of people at risk?

[30] Wofford (2022); Birnbaum (2021); Mac and Silverman (2021); Horwitz and Seetharaman (2020); Seetharaman (2018). See Chapter 4 for more details.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

We might be more inclined to label a governance failure as a failure of will when we think that there are more normatively troubling conflicts of interest playing a large role. For example, Amazon makes profits even on counterfeited goods (so long as they go undetected further down the chain); Facebook users excited by political misinformation may stay on the platform longer and look at more advertisements. If companies fail to control profitable kinds of conduct creating third-party harms, we may be inclined to demand a higher degree of corporate economic burden before we treat that failure to control as a case of inability rather than lack of motivation.

By the same token, however, we ought to recognize that, as noted above, in many situations the economic incentives of platforms may give them strong reasons to effectively regulate their users. If Facebook allows itself to be turned into 4chan, then only the sorts of people who hang out on 4chan will go there – a much smaller and more pathological user base who cannot support a megabillion dollar company. And some of our major companies have actually felt significant economic bites from governance failures: YouTube had an exodus of advertisers in 2017 due to a "brand safety" scandal after major companies found their products advertised on extremist videos (Solon 2017); Nike and Birkenstocks stopped selling their products directly on Amazon due to counterfeiting (Bain 2019). As of this writing, Elon Musk's erratic leadership of Twitter is creating comically extreme brand safety threats for advertisers – shortly after he acquired the company, he permitted anyone to purchase a "verified" checkmark for $8, and verified accounts shortly came into being parodying numerous major corporations. The most grimly amusing example: Someone bought a checkmark and then, under the name "Eli Lilly," falsely (alas) declared that insulin would be given away for free. Unsurprisingly, the company canceled its Twitter advertising (Harwell 2022).

If a company is experiencing those kinds of consequences and nonetheless fails to control the behavior that leads to them, it gives us some reason to interpret those failures as rooted in inability, due, for example, to technical difficulty or to divergent incentives between the top-level leaders whose intentions animate the company's goals and lower-level employees implementing those intentions (on which more in Chapter 4). For example, two years after many advertisers announced a boycott of Facebook in protest of its hosting and profiting from white supremacist content, it still – company representatives say inadvertently – serves advertisements against such content and in some cases algorithmically generates white supremacist pages based on user interests (Nix 2022). Similarly, in September 2022, Twitter (pre-Elon!) discovered that it had run advertisements from several major companies like Coca-Cola on accounts full of child sexual abuse material (CSAM) (Fingas 2022).[31] In view of the fact that such content violates both companies' stated policies, has

[31] Incidentally, this also illustrates the extreme difficulty of automated enforcement, since there are vast databases of CSAM which companies use to detect image "fingerprints" through the PhotoDNA program – but that isn't enough to keep it reliably off Twitter.

led to financial consequences from advertisers, and in the case of Twitter's CSAM problem is also a universally loathed major crime throughout the planet, it's pretty tempting to see the failure to eliminate it as "inability" rather than "unwillingness," at least relative to existing levels of company investment in enforcement.

Our choice of whether to label a company's governance failures as a case of inability or unwillingness, in turn, may implicate the strategies we choose as democratic citizens operating governments to remediate them. In the case of inability, we have some reason to prefer interventions on platform features like their underlying corporate governance or employee relations, or the technologies available to them (e.g., by organizing licensing schemes for such technologies). By contrast, in cases of unwillingness, we have reason to prefer more coercive interventions such as the threat of fines and antitrust action. However, I shall suggest in this book that there are some interventions that can actually avoid the inability/unwillingness dichotomy, insofar as they can remediate both problems at once, that is, to give platforms both the ability and the incentive to govern their users adequately.

It is worth identifying that there is, of course, an unavoidable amount of disagreement and contestation on what adequate governance looks like. The example of right-wing misinformation on social media could serve again: Many on the political right would deny that much of the material in question actually does misinform people. Regardless of the proportion of the debatable material that is removed, those on the left will always think that it's not enough, those on the right too much. Still, this project does not require us to resolve such questions. One of the tasks of any effort to build governing capacity includes building the capacity to come to reasonable, even if imperfect, resolutions of highly controversial cases.

The reader may object that if adequate governance can't be observed, then we can't tell whether any program to develop it is a success. To this, my response is twofold. First, we can observe improvements in uncontroversially terrible user behavior, such as despotic militaries inciting genocide and Russian spies pretending to be Black Lives Matter activists on social media, or counterfeit products on Amazon. Second, we can observe relatively uncontroversial process improvements, such as the inclusion of minorities of all stripes (racial, religious, ethnic, etc., relative to a country or a problem, for example). Given where we stand today with the immense number of social harms created by platforms, we should work to solve the easy (in an evaluative sense) cases before worrying about the hard ones.

SCHOLARS HAVE THE TOOLS TO IMPROVE PLATFORM GOVERNANCE: BORROWING FROM POLITICAL GOVERNANCE

In the tradition of academic books, part of my mission is to fill a surprising gap in the scholarly literature. There are, of course, countless scholars writing about platform governance, from disciplines such as law (one of my home fields), communications,

science and technology studies, and the like. And there are numerous scholars in political science (my other home field) and allied fields such as political economy writing about the problems associated with platforms (e.g., Tucker et al. 2018; Zhuravskaya, Petrova, and Enikolopov 2020). However, the political science literature mostly resides in the behavioral side of the discipline, that is, from scholars who empirically study how people participate in political activity.

Yet there's another side to political science, traditionally known as "institutions." Scholars in that half of the discipline, in conversation with allied disciplines like economics, sociology, and history – as well as more applied disciplines such as public policy, urban planning, and ecology – write about the effects of different organizational forms and patterns of interaction on aggregate behavior. Such scholars consider questions such as the conditions under which it might be possible to bind people (particularly, but not exclusively, top-level officials) to complying with legal rules (e.g., Hadfield and Weingast 2014; De Lara, Greif, and Jha 2008), how the independence of judges is preserved (e.g., Ferejohn 1999), the relationship between different types of property rights and the ability to manage shared resources (e.g., Ostrom 2003), and the like.[32]

There is a dearth of work from the institutional tradition of political science and its allied disciplines noted above on the problem of platform governance itself. While there are a handful of articles in the vein about specific problems and specific platforms, there is no comprehensive or book-length treatment.[33] This book aims to begin the conversation on that broader basis.[34]

[32] Fascinatingly, there is a surprising affinity between some of the modern governance literature in the disciplines I have described and a more technologically oriented and science-fiction sounding discipline with essentially no direct intersection with political science, namely, "cybernetics" (Beer 2002). More or less, as far as I can discern, the cybernetics people talk to ecologists and complexity theorists, and then complexity theorists and ecologists talk to political scientists (mostly thanks to Ostrom). And the technology people sometimes talk to the cybernetics people because the name sounds like something out of science fiction. Thus, the bizarre sociology of the intelligentsia. For one story about the intersection between cybernetics and the kinds of organizational political theory that partly animate this book through the lens of anarchist(!) philosophy, see Swann (2018). Some of the ideas described below, such as on the ineffectiveness of rigidly centralized control, also feature in the cybernetic literature (Swann 2018, 433).

[33] Some examples of the most important article-length work setting up the foundations of this nascent literature from the political science and allied side include Gorwa (2019b, 2019a), Napoli (2014, 2015), Srivastava (2021), and Caplan and boyd (2018).

[34] This approach is also self-consciously a product of my own unusual intellectual location. There is an active platform governance literature which I consume (although given the volume and rate at which scholarship in platform governance is produced, I've probably missed important work), but in which I have not previously been a participant. There is a sense in which I intrude on that conversation as an outsider – as a political theorist and constitutional scholar drawing on external fields to intervene in an existing literature, with all the advantages (cross-pollination of ideas, a fresh perspective) as well as the disadvantages (the risk of repeating or missing ideas extant in that community, the potential misuse of internally generated terms of art) this entails. At the same time, I am unusually familiar with the problem space from an odd angle: As described in the Appendix to this Introduction, I have worked within one of the most major platform companies on initiatives to address some of its most important

This book rests heavily on a broad cluster of theoretical ideas that has crossed many disciplines and been associated with a number of prominent scholars. The lodestar points of this cluster include, among other things, Ostrom's (2015, 2010a, 2010b) work on polycentric governance, Hayek's (1945) on the problem of knowledge, Scott (2008) on high modernism, Jacobs (1992) on urbanism, Dewey's (1927) democratic experimentalism, Ober's (2008) historical work on democracy and knowledge in classical Athens, and a mass of work in public administration and related fields that often goes under the name "New Governance" associated, for example, with Mark Bevir (2013).[35] This seemingly diverse set of ideas tends to converge, for different reasons, on propositions such as the following:

- Centralized top-down command-giving is often ineffective because of its difficulties with integrating knowledge from the periphery and offering legitimate rules to diverse constituencies.
- Many effective institutions of governance are grown or evolved out of the immanent behavior of people trying to solve their own problems, rather than designed or imposed.
- Novel governance strategies can be developed by permitting some play in the space between means and ends, for example, by creating local sub-institutions empowered to develop experimental or even idiosyncratic techniques to pursue shared goals in the context of dense cross-institution communication and learning.
- Rules and governing institutions frequently require revision in the light of practical experience with their operation.
- Agents and organizations engaged in the activities of governing can often be more effective when organized into complex structures including features such as overlapping and multi-scale jurisdictions and collaborative networked relationships drawing on markets and informal social interactions.

Yet despite the challenges to centralized governance and the empirical successes of alternative forms represented by this literature, the major platform companies uniformly have a centralized, top-down, authoritative governance structure for user behavior. With very few exceptions (mostly Reddit, Discord, Wikipedia,[36] and to a

governance challenges, and maintain a close and active engagement with workers and former workers on these problems from numerous companies through a nonprofit organization (the Integrity Institute) in which I have had a high degree of involvement. I am also a longstanding participant in the conversation on governance more generally through the central normative construct of the "rule of law," on which I have previously published two books and numerous articles. So I am, somewhat bizarrely, an outsider to the academic literature on platform governance but not to the problem of platform governance in actual implementation nor to the problem of governance in the abstract. I ask the reader to consider the arguments offered in this volume in that light, and excuse any distortions which I inadvertently impose on existing scholarship in the more narrow "platform governance" academic field.

[35] On the relationship between Dewey and new governance, see Simon (2010).

[36] While I mention Wikipedia in several places for the purposes of comparison to some of its governing entities, it does not meet the definition of a platform as used in this volume.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

limited extent the Meta Oversight Board), the rules are made in a corporate hierarchy in somewhere like Menlo Park or Seattle, and enforced by a combination of machine learning algorithms and human enforcers directly answerable to the corporate chain of command and nobody else. And this is so even though the major platforms are managing globe-spanning user conduct, with deeply interconnected networks of people generating complex emergent patterns of behavior in a context of extreme diversity – quite possibly the least suitable setting for the centralized command-and-control style.

The enterprise of this book is also inspired by (although does not as directly deploy) intellectual frameworks that have long recognized that a higher-level abstraction can be used to analyze both states and firms, namely, the organization. Some of the foundational work in the political institutions research program is built on the recognition that business companies and political entities face similar organizational and governance problems, and consciously applies economic theories like Coase and Williamson's theories of the firm (e.g., Williamson 2005) to political states (e.g., Weingast and Marshall 1988; Moe 1984; North and Weingast 1989). On the other end of the social sciences, organizational theorists in sociology (W. R. Scott 2004) such as March, Olsen, Powell, and DiMaggio (e.g., March and Olsen 1984) have identified that ideas like diffusion of strategies and logics of legitimacy and appropriateness apply across organizational contexts. Early work in the new economics of organization specifically attends to governance as an enterprise that can take different structural forms – most famously understanding firms and markets as alternative ways of arranging transactions (Williamson 1996, 133).

As yet, efforts to apply these theoretical frameworks to platforms are in their early stages. The most interesting exception is Marxist economist Laurent Baronian's (2020) effort to conceptualize platforms' relationships with "users" (conceptually centered on, but not limited to, transactional platforms' relationships with worker-users, as with Uber drivers) as novel solutions to the management problem of determining the boundaries between firm and market. But I contend we can learn more by explicitly drawing from the application of these theoretical frameworks not just to firms but also to states.

The discussion in this volume will also be guided by both the strategic and the normative. The governance literature, like most contemporary political science and economics, tends to be focused on the management of the strategic incentives of participants (here, companies, users, and governments) as well as the structural features of an interactive environment that make it possible for participants to respond to those incentives (e.g., the sharing of information, and second-order incentives to conduct that sharing, and so forth). But the evaluation of the predicted outcomes of those incentives, and thus any recommendations in terms of actual policy or design outputs, necessarily depends on external normative standards.

The core normative presuppositions of this book can be described in terms of the concepts of democratization and inclusion. That is (very briefly, and with elaboration spread through the rest of the pages that follow), I suppose that (1) people ought to be able to run their own lives, collectively, through regulatory institutions that are accountable to the regulated,[37] and (2) we should be alert to the way in which governing arrangements can go wrong by failing to identify the people who should rightly be included. The latter category, for example, includes the danger that those in the so-called "developed" world will economically and politically dominate those in the "developing" world in ways that are objectionable both because they undermine the self-governance of the latter and because they represent the continuing unjust legacy of historical conquest and colonization. But that category also includes existing biases and exclusions within countries that company governance might replicate, such as the exclusion of racial, religious, gender, and cultural minorities.

There's a close affinity between this book's democratic and inclusive normative side and its pragmatic focus on modern theories of governance. I shall argue that the best path forward for democratizing platform governance and for including the legitimate claims of those who are not the powerful elites from the wealthy developed world in the decision-making processes involves the creation of institutions that can incorporate ordinary people into polycentric and densely interconnected governance processes much like those described by modern governance scholarship (but with a broader popular element than is traditional for the sort of "new governance" that mostly runs together NGOs, governments, and corporations). This is not merely coincidental: In a highly diverse ecosystem operating at an immense scale, like every major platform, the imperatives of (normative) legitimacy and the imperatives of effectiveness come together, for both demand the deep-down inclusion of a wide and ever-expanding variety of knowers and stakeholders.

A final key influence that requires specific discussion in the Introduction – for it has guided the entire book in a variety of subtle or explicit ways – is John Dewey's major work of political theory, *The Public and Its Problems*. Dewey begins, as the title suggests, by giving an account of the domain of the public, namely those interactions between people which have an effect on third parties (what contemporary economists like to call "externalities"). The nature of the state, he argues, cannot be discovered by some theoretical derivation from first principles but rather arises out of the efforts to solve these third-party effects – and thus the nature of the state is different under different social, technological, and economic conditions. Moreover,

[37] This one-clause summary is self-consciously neutral with respect to whether the implication of this idea in the platform context is that democratically elected governments ought to control platforms or their users, or that people ought to have democratic control over those platforms directly – as it will turn out, the answer I offer is somewhat more complex, and will require some development.

this suggests that there is an experimental quality to the organization of states and that the goal of the study of politics is to help build the conditions under which experimentation and learning can be successful.

The Deweyian approach to the state (a kind of pragmatist Hegelianism) is perhaps best summarized in the following passage, worth quoting at length:

> In no two ages or places is there the same public. Conditions make the consequences of associated action and the knowledge of them different. In addition the means by which a public can determine the government to serve its interests vary. Only formally can we say what the best state would be. In concrete fact, in actual and concrete organization and structure, there is no form of state which can be said to be the best: not at least till history is ended, and one can survey all its varied forms. The formation of states must be an experimental process. The trial process may go on with diverse degrees of blindness and accident, and at the cost of unregulated procedures of cut and try, of fumbling and groping, without insight into what men are after or clear knowledge of a good state even when it is achieved. Or it may proceed more intelligently, because guided by knowledge of the conditions which must be fulfilled. But it is still experimental. And since conditions of action and of inquiry and knowledge are always changing, the experiment must always be retried; the State must always be rediscovered. Except, once more, in formal statement of conditions to be met, we have no idea what history may still bring forth. It is not the business of political philosophy and science to determine what the state in general should or must be. What they may do is to aid in creation of methods such that experimentation may go on less blindly, less at the mercy of accident, more intelligently, so that men may learn from their errors and profit by their successes.[38]

This perspective seems entirely sound to me when confronting novel forms of governance with a novel underlying "public" and trying to figure out what to do about it. While Dewey assumed that states would be geographically contiguous (Dewey 1927, 39–43, 212–13 – he was, after all, writing almost a century ago), I think that he would find the notion that platforms have their own publics, and thus are at least in the same family as states to be fairly congenial.[39] He would probably agree that the process of figuring out how to govern externalities generated by activities in these novel, somewhat state-like, entities is at heart experimental, and that what we must do is aim to build the conditions in which that experimentation can be carried out. I argue that ultimately those conditions include a kind of radical inclusiveness that recognizes that the Deweyian public for platforms is global, and that it is this global public which must be permitted to experiment. I further argue that even the scope of inclusion, that is, who is to be involved in running the institutions of platform governance itself, is itself ineluctably experimental.

[38] Dewey (1927, 33–34).

[39] He did recognize that political boundaries could diverge from the publics constituted by the effects of economic activity and change in communication technologies. See ibid., 107, 114.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

Because demands for inclusion are likely to come from unanticipated directions, a chief design criterion for platform governance will be to build institutions capable of responding appropriately to those demands, and thus learning from contestation even over the scope of its own stakeholders. This is all, to my mind, deeply Deweyian and Dewey's influence runs throughout this book.

WHERE WE'RE GOING

The remainder of this book is divided into three substantive parts. The first part, consisting of Chapters 1 and 2, further develops the general approach in the Introduction. Chapter 1 more carefully defines the platforms under consideration and describes their general characteristics.

Chapter 2 draws the analogy that drives the approach of this book between platforms and states – in particular, failed states, which lack the ability or the incentive to adequately govern those who use their services. It addresses several major objections to the enterprise of adopting a capacity-building approach to the project of platform governance, where our political states work to shape the incentives and abilities of platforms in order to govern their users.

Part II, consisting of Chapters 3 and 4, returns to the problems that lead this Introduction. Chapter 3 begins with the Myanmar genocide and argues that Facebook's culpability in that genocide resulted from a characteristic problem, also experienced by governments, of bringing knowledge from the periphery of the governed domain to the center. The chapter argues that democratic institutions, organized in ways attentive to the dispersal of authority and the aggregation of information across space and scale, can mitigate such knowledge problems.

Chapter 4, in turn, takes up the problem of Donald Trump – both the propaganda that led to his election and the companies' seeming inability to control his supporters while in office – leading up to, but not limited to, the January 6 coup attempt. It contextualizes these events in a broader narrative about social media "political bias," which I interpret less as a genuine problem of bias and more as an effort by politicians to intimidate companies into under-moderating their sides. Such efforts aim to leverage the inability of the companies to exercise self-control in the face of short-term temptations and threats. Platforms, like governments, have problems of internal governance, in which personnel have incentives that diverge from the interests of the overall organization or operate under suboptimal time horizons. In the literature on governments, the tools to mitigate these problems tend to travel under the rubric of "the rule of law." I defend the idea of dispersing power to independent institutions under the control or at least supervision of diverse groups of employees and non-employees as a key tool to create a kind of platform rule of law.

Ultimately, the two chapters of Part II point toward the same primary ideas. To wit: the social organization of governance matters; effective governance requires that

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

people with knowledge (about what is happening, about their needs) and distinctive interests be assembled in network structures where they have an incentive to share their knowledge – incentive conferred in substantial part by genuine power over outcomes that matter to them and the capacity and incentive to negotiate over the use of shared resources. Doing so ultimately requires the conscious building of inclusive processes in which currently under-represented stakeholders, such as those in the global South, as well as those who have the latent power to control platform companies, such as their own workers (broadly understood) are organized into groups with overlapping authority over key governance decisions. The title of this book – *The Networked Leviathan* – thus has a double meaning, referring first to the existing nature of platform companies, which occupy quasi-governmental roles due to their leveraging of network effects; but also referring to the capacity of interventions on the network structure of platform users and workers (and secondarily governments and civil society) to create a new kind of governing structure. The title is also ironic, for, *contra* Hobbes, I ultimately argue that Leviathan must share rather than hoard his power.

Finally, Part III, consisting of Chapters 5 and 6 as well as the Conclusion, turns to practical implementation of the book's overall approach. Chapter 5 examines the design and performance of the most prominent recent innovation in platform governance, the Facebook (now Meta) Oversight Board.[40] It contextualizes the examination of that board in the rule of law ideas of the preceding chapter. In the course of this analysis, it also develops some ideas about the normative function of platform legalism in building a kind of platform identity which may be valuable in resolving controversial governance issues.

Chapter 6 makes concrete proposals for the design of polycentric, decolonial, democratic governance institutions in the platform economy. Chapter 6 focuses on institutions that platforms could, in principle, build themselves; it could be understood in the first instance as being directed at senior platform executives. The end of Chapter 6 and then the Conclusion address governments, describing some ideas for ways they (particularly, but not exclusively, the United States and the European Union) could give companies the incentives necessary to implement some of these reforms, as well as other direct beneficial interventions that states could make on platform governance.

APPENDIX TO INTRODUCTION: ADDRESSING SOME ETHICAL CHALLENGES

Before turning to the substance, I must disclose and address some potential issues of research ethics. I do not write as a fully independent observer. Most importantly, I've taken money from Facebook to apply my academic knowledge on areas closely

[40] I had some involvement in the design of the board; see the Appendix to this Introduction for a full disclosure and description.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

related to the topic of this book. From the summer of 2018 through the end of 2019, I was a paid consultant/contractor for Facebook: For several months in 2018, I was embedded full time in the company's Civic Integrity team advising on considerations of democratic theory and ethics relating primarily to elections; thereafter, I continued to consult on a part-time basis with that team as well as with Facebook's Product Policy Research team supporting the design of the Facebook Oversight Board. In the latter capacity, I co-wrote a report with Radha Iyengar Plumb, then Facebook's Head of Product Policy Research, which Facebook distributed in conjunction with an announcement about the Board.[41]

In addition to paid work, I also have other interactions with the technology industry that arguably might undermine my research objectivity. I've given informal (and unpaid) consultations to staff from other prominent technology companies who have sought my advice about their governance institutions. I've also been a participant in industry conversations, including conversations under confidentiality obligations meant to facilitate frank conversation. (This paragraph uses the nonspecific plural largely to preserve a degree of confidentiality in all of the interactions it describes.) Finally, I've been involved with research and advocacy organizations seeking to produce reforms in how social media companies do content moderation. I am a founding fellow of the Integrity Institute, a nonprofit organization composed of present and former "integrity workers" at social media companies dedicated to advising policymakers, press, and public on the challenges of platform integrity; I am also a member of the Integrity Institute's nonprofit (corporate) board as well as its Community Advisory Board.

These activities might reasonably raise several concerns in a reader of this book. First, I might be compromised by past financial gain or the anticipation of future financial gain: The arguments in this book might be unconsciously or even consciously influenced by a pro-company or just pro-Meta bias as a result of my financial interests. This is potentially particularly problematic because the company that paid me was Facebook, which has a record of getting journalists and academics into compromising financial entanglements.[42]

Second, confidentiality obligations that I've accepted, either formally – my work at Facebook was done under an industry-standard nondisclosure agreement – or

[41] Paul Gowder and Radha Iyengar Plumb, "Oversight of Deliberative Decision-making: An Analysis of Public and Private Oversight Models Worldwide," Appendix E to "Global Feedback and Input on the Facebook Oversight Board for Content Decisions," released June 27, 2019, https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf#page=138.

[42] For example, *New York Times* columnist David Brooks recently failed to disclose to his editors that he was producing a corporate blog post including a paean to the wonders of Facebook Groups for the company, to promote a study by researchers at NYU on how wonderful they were; while Brooks apparently was not directly paid for this work, something called the "Weave Project," which he founded, at the Aspen Institute, had received funding from Facebook, as did the NYU researchers whose report Brooks's blog post for the company was meant to introduce (Silverman and Mac 2021).

informally, such as under "Chatham House Rule" norms in various conversations in which I've participated, might compromise my ability to fully use knowledge that I've gained in these relationships to improve the scholarship represented by this book. From the other direction, my work in this book might violate those confidentiality commitments.

Third, I may have a bias as a result of the tasks I have done and the social contexts in which I'm embedded – I associate with many past and present platform company employees and directly worked on some of the things discussed in the following pages. Nobody wants to come to the conclusion that their own efforts were wasted, or the things that they and their friends have worked on are harmful. Self-justification bias is a powerful force.[43]

To some extent, these problems go along with the territory.[44] Without having relationships with companies and others associated with the industry, I wouldn't have the knowledge necessary to write this book sensibly – one of the most unfortunate aspects of the current debate on platform regulation is the extent to which it's dominated by, on one side, critics of the industry with no information as to the actual constraints companies face; on the other, industry insiders with expert knowledge of those constraints but tainted motivations. I am trying to straddle that gap, but I may not be successful.

In an effort to ameliorate these problems, I've taken the following steps. First, of course, is this Appendix: I've endeavored to write a full disclosure that communicates all relevant facts, provided early on in the main text of the book rather than buried in a footnote or at the end.

Second, I have taken steps to sever my personal financial interests from those of the companies under discussion. As of the sending of this book to press, I have not done any paid work for Meta or any other technology company since the end of 2019, and I do not have any offers, ongoing discussions, or plans to do any in the future. (However, I have not committed to refusing any such work that may come along after this book is published.) Nor have I or any institution with which I'm affiliated received any research support for the production of this book from any for-profit company; the only financial support outside of Northwestern University's ordinary resources that contributed to this book was a grant from the Knight Foundation. Similarly, I have, to my knowledge, no discrete investments in any technology company, although I do have mutual fund investments over large portions of US and

[43] I think this is the least important of the risks. The place where it is most likely to crop up is in a bias toward believing that the Meta Oversight Board is a good idea, but, as you will see in Chapter 5, I have no qualms about being critical of the Board's design where warranted. Moreover, part of the reason I agreed to work on the Oversight Board was because I think it's fundamentally a good idea – as evidenced by the fact that I published scholarship in favor of the general idea of independent judges long before I had any affiliation with Facebook (e.g., Gowder 2014b, 2016).

[44] Tarleton Gillespie (2022, 2–3) has a "methodological note" in a recent article of his which is particularly thoughtful on these kinds of issues.

international markets, including the technology sector, in my personal investment and retirement accounts, which are managed by professional fund managers without my input in choosing individual stocks.[45]

Third, on the informational side, no company has reviewed or approved the manuscript for this book. Because I am covered by a nondisclosure agreement relating to all work done for Meta, I have been very careful to identify publicly reported sources for any information about Meta discussed in this book.[46]

[45] However, I acknowledge that this does not completely free me from potential financial bias. My views might be biased by the hope of *future* financial opportunities from platform companies. But that's true of anyone who writes academic work about a topic in which wealthy companies are also interested; there's nothing special about what I've done in the past that changes this risk.

[46] In practice, when I have discussed more sensitive or controversial issues involving Meta this has amounted to citing many news articles drawn from the so-called "Facebook Files," documents leaked by whistleblower Frances Haugen. I do not assert the truth or falsity of anything reported in any of these sources, which should stand or fall on their own.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

1

The Nature and Problems of Platforms

Since this is a book about platform governance, it should start by saying what a platform is. As noted in the Introduction, the things that go by the name "platform" probably bear a family resemblance to one another rather than a rigid set of defining and distinguishing features. Accordingly, rather than attempt a formal definition, I propose to focus on the qualities that seem to make arrangements of humanity such as Facebook and Twitter so difficult to govern. Thus, I will use the term "platform" for a unified set of human-usable tools or services on the Internet that enables people not affiliated with the company that operates the tools to interact with one another, typically monetizes that interaction, and is characterized, generally, by the following properties (all of which are matters of degree rather than binaries):

1. It provides positive *network externalities*, that is, users benefit the more other users are present.
2. It provides strong *network affordances* for users to interact with distant (either in geographic or social terms) others.[1]
3. It operates at a *substantial scale* as a consequence of the incentives that its network character gives it.
4. It offers affordances for and is actually characterized by a *substantial diversity of uses and norms* as a consequence of the scale and network affordances (i.e., with lots of people who are different from one another, they find lots of things to do with the tool).

I'll fill out each of these ideas below, with some diversions and implications.[2]

It will often not be possible to decree, as a simple up or down determination, whether any single service is or is not a platform, but it will be possible to at least

[1] The notion of social distance captures the way that even a geographic platform might create more dense geographically bound social networks – if I join Nextdoor, for example, I can interact with many more of my neighbors than those whom I might ordinarily see.

[2] In the existing literature, my definition (such as it is) is perhaps closest to that of Srnicek (2016, 43–48). An earlier, related use of the term "platform" in the technology industry context centered on the capacity of other applications to run on a service, as with the idea of a mobile phone with an

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

identify differences in platform-ness across different internet services with reference to the above properties.[3] For practical purposes, I will treat the creation or facilitation of significant behavioral novelty among users as a key indicator of the platform character of some service. The reason for this is pragmatic: While non-platform types of intermediaries also find themselves engaging in a certain amount of behavioral governance (e.g., payment processors like Stripe or the credit card networks regulate the sorts of goods and services their users may sell, such as by refusing to serve pornographers), the governance challenges of such intermediaries aren't nearly as interesting. They're basically just facilitating the same kinds of transactions that banks have facilitated for a very long time and can probably rely on similar regulatory strategies.[4]

Ultimately, I agree with Lobel (2016, 101–2) who identifies this as part of the nature of the domain: because the core of what platforms offer is a set of affordances

app store as a "platform" or the early market positioning of Facebook as a place to make applications, such as the infamous game "Farmville," available to its users. That alternative usage of the term is equivalent to the "computational" sense of platform described by Gillespie (2010, 349). This volume is not primarily concerned with "platforms" in the computational sense. Rather, the sense that seems most relevant to the current analysis is what Gillespie describes as "figurative," in which one uses the word "platform" to describe a collection of affordances allowing one to do something – in which we say, for example, that the president of the United States has a large platform in virtue of his or her capacity to reach a broad audience. Another helpful distinction in the existing literature is between "transaction" and "innovation" platforms (Cusumano, Gawer, and Yoffie 2019). Here, I focus entirely on "transaction" platforms.

[3] Another important question, with which this book cannot wrestle, is the distinction between the application layer and the infrastructure layer: When companies like Cloudflare decide to control the behavior of the companies (who themselves host users) who use their services (cf. Zittrain 2021), is this different from when Facebook does it? I am inclined to draw a fairly sharp line between the two in view of the fact that most infrastructure companies lack the network properties of what I am calling platforms, but this may be controversial. Relatedly, it is important to note that not all internet services that engage in the enterprise of "content moderation" are platforms, and the same service may function as a platform in one context but not in another. For example, workplace communication tools such as Slack and Basecamp may need to engage in content moderation, however, to the extent they tend to be deployed in closed universe settings on top of an existing physical network (such as within a team at a specific employer), they are unlikely to feature the same kind of diversity as something like Twitter. However, if they are deployed more broadly, for example, an open Slack instance for everyone in the world with a specific set of interests, they may function as platforms. Moreover, even in closed-universe contexts, companies conducting content moderation can experience problems similar to those described in this book, so the analysis herein may be of some use to them. For an amusing example: It turns out that at least one corporate communications service provider moderates the word "bone," doubtless because of its use as sexual slang and the problem of workplace sexual harassment. But that didn't work so well when a society of fossil-studying paleontologists subscribed to the service! Maria Cramer, "Paleontologists See Stars as Software Bleeps Scientific Terms," *New York Times*, October 18, 2020, www.nytimes.com/2020/10/18/science/paleontology-banned-words-convey.html.

[4] Even this attempt at line-drawing can break down: A company like Uber is a quintessential platform, but it is debatable whether it has created behavioral novelty to the degree of something like YouTube – perhaps it has, in view of the way it has blurred the lines between commercial and individual drivers, or perhaps its primary novelty is related to the scale of its operations. Oddly, these definitional boundaries can also blur in the other direction. Content creators who use YouTube can also treat it like the Uber sort of "gig economy" arrangement to the extent they rely on payments from YouTube's monetization program (Caplan and Gillespie 2020, 2).

to connect users, and because those affordances are compatible with a huge array of economic (and noneconomic or semi-economic) activities, in the context of cultural as well as economic incentives for continual innovation, platforms will necessarily resist firm definitional line-drawing.[5]

1.1 THE CORE FEATURES OF PLATFORMS

1.1.1 *Positive Network Externalities*

The presence of positive network externalities is perhaps the most basic and well-understood feature of platforms. If I'm a member of a social media site, I typically get more value from that website to the extent that people I already know – or people I want to know – are also using it; likewise, if I'm either a seller or a buyer on transactional platforms like eBay, I get more value to the extent that there are more potential counterparties in the marketplace. This doesn't require much elaboration because it is so well known, but subsequent sections will fill out some of the related details.

1.1.2 *Interactive Network Affordances*

Platforms' core function from the user standpoint is to provide a system to enable a large number of people to engage in various kinds of interactions with one another. For convenience, we can call this central feature of a platform its organization around "interactive network affordances." Those interactions might be economic (eBay, Etsy), interpersonal (Facebook, Twitter, YouTube), or both (LinkedIn).

Much of the added value of any platform is its capacity to expand the effective scope of a network that a given user can access – at least potentially. If I have enough followers on Twitter, I can reach many more people with my hot takes on the Supreme Court than if I just speak to my friends. If I have something to sell, I can reach people all over the world on eBay rather than simply relying on my own city. Even in platforms nominally rooted in a user's own social network (Facebook) or in specific geographic locations (Airbnb), network expansion is a key affordance (interacting with friends of friends, discovering otherwise-unadvertised rental properties).[6]

[5] That being said, many of the challenges and techniques of platform governance are also challenges and techniques of intermediary governance more generally. The banking system provides numerous examples of, for example, the use by governments of private intermediaries to regulate their users (as with anti-money-laundering legislation that requires banks to report certain kinds of transactions to the government) – a relationship between government and company that might not be seen as meaningfully different from the ones that show up in major internet platform contexts with legislation such as the Digital Millennium Copyright Act.

[6] For a discussion of how one social network failed, at least within a specific context, in part because it could not deliver network novelty to its users, see Pearson and Moskal (2017, 122–26). For a useful early review of some of the economic concepts, see Evans and Schmalensee (2010).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

"Expansion" is probably not the correct term in all cases, for human cognitive and time limitations continue to be a constraint on what we might call the effective size of a person's addressable network. I can only consider so many lodging options or regularly interact (except in a broadcast mode) with so many other people. Another key aspect of the network affordances that platforms tend to offer, rather than raw expansion, is customization of one's network (thanks in part to recommender algorithms, discussed below), which tends to allow people to find niches not otherwise available to them.[7]

This network niche-finding capacity can have dramatic consequences. Consider the phenomenon of meme stocks: A subreddit called "Wall Street Bets" became, for mysterious reasons, a kind of focal point for people who not only traded stock recreationally, but for those who were interested in the pure entertainment value of a kind of expressive stock buying – choosing, from some combination of nostalgia, nihilism, and perhaps a (successful) effort at market manipulation, to aggressively buy companies like GameStop. It turns out that lots of people were watching, and at least some of those watchers may have recognized the power of this group to affect pricing in otherwise lightly traded stocks. Elon Musk joined in.[8] Accordingly, lots of those watchers eventually bought in themselves on the anticipation or observation of price movements, and GameStop and other "meme stock" companies had massive rallies. There are lots of reasons for the rise of the meme stock, including the phenomenon of virality, discussed in detail below, but certainly one significant part of it is that Reddit created a global as well as publicly visible place for people who were both interested in stock trading and a little bit nihilistic (r/wallstreetbets bills itself as "Like 4chan found a Bloomberg Terminal").

Social media platforms are particularly novel because they tend to afford users the opportunity for multilateral interactions. That is, users may interact with a large number of other users at once. For example, on social media, users in their capacity as content producers may potentially reach an unlimited number of others both sequentially (as with "retweeting") and simultaneously (as when a user with a lot of followers has broad reach for a single piece of content). Similarly, a user in his or her capacity as content consumer may consume content from a large number of others, as by following many people on Twitter. This quality of social media derives in part from the fact that (unlike the sale of physical products by transactional platforms) the things (content) that users supply to one another are non-rivalrous, and hence suitable for resharing (and with it distinctive problems like virality) or for bulk

[7] For example: I own far too many Apple Watch bands (there are lots of them; it's kind of a fashion accessory). It turns out there are several subreddits and a Discord chat that allow me to talk to other people who also own far too many of them, trade extras, find deals off the (outrageous) face price, and so forth – reminiscent of the baseball card conventions of my youth in the 1980s and 1990s, except permanently available everywhere. (You might reasonably question the social value of this.)

[8] By posting a link to r/WallStreetBets in a tweet with the text "Gamestonk!!" (https://twitter.com/elonmusk/status/1354174279894642703, January 26, 2021).

consumption and second-order interaction (such as Facebook or YouTube comments). Indeed, we can identify social-media-like phenomena (including potential governance problems) in seemingly nonsocial platforms by observing that users primarily interact with one another by exchanging non-rivalrous content. For example, the open-source software code-sharing website GitHub, though very different from what we think of as "social media," could display similar network dynamics and cause similar social problems as the social media platforms.[9]

1.1.3 *Skim as a Revenue Model and Incentives to Scale (and Maybe Concentration)*

Commercial platforms primarily make money by skimming from the top of the activity that they facilitate. This is most obvious in the case of platforms oriented around economic activity, which typically take an explicit commission. But it is also true of platforms oriented around social activity, which skim some of the attention of their users off of the top and sell it. For the most obvious (and, to a professor who sometimes tries to show videos in class, deeply annoying) example, YouTube requires its users to devote some of their attention not to content created by their fellow users but to advertisements.

The skim-based revenue model contributes to strong platform incentives to scale – to increase users – as revenue is increasing in number of active users.[10] Moreover, it also gives platforms economic incentives to promote increased activity by existing users ("engagement" in the social media context).

The prominence of positive network externalities in the incentive structure surrounding platforms implies that they tend to scale organically: for both companies and users, the benefit received from a platform is increasing (only up to a point for users – see Section 1.1.4 for recommender algorithms) with the size of the platform's userbase, so platform growth tends to be self-reinforcing. Successful social media platforms in particular tend to become very large as they devour existing physical-world-social networks (i.e., everyone can find their friends there), or they tend to fail quickly. A social platform can also collapse midstream via cascade processes as users depart – this is what happened to the early social network Friendster (Garcia, Mavrodiev, and Schweitzer 2013). Of course, every company wants to grow, but the incentives are much stronger for growth in platform companies than for other types of business, which have other ways to be profitable and aren't in danger of such failure cascades.

[9] Indeed, to some extent, code-sharing platforms have already evidenced some of the characteristic problems of social media, such as quasi-viral content that turns out to be malicious, in the form of security holes introduced by bad actors into popular open source libraries (e.g., Goodin 2018).

[10] Meta is perhaps the most famous example of this imperative for growth, which has been the subject of frequent discussion in the press (e.g., Roose, Isaac, and Frenkel 2020; Kantrowitz 2019; Hao 2021). This is doubtless because of Meta's particularly prominent acquisitions of Instagram and WhatsApp, which – especially the latter – are most understandable as investments in new groups of active users.

In practice this may also imply concentration because using platforms is costly (if only in time): If I have to search Amazon and Walmart and eBay and Mercari every time I shop for a product, I'm likely to want to abandon the smaller of these platforms as soon as practicable in favor of a platform that is more likely to have everything I need in one place.[11] This may imply a winner-take-all platform economy even aside from any anticompetitive behavior by companies, although companies may also promote growth in part to fend off competitors by ensuring that users who are economizing on on-platform time to meet some need which scales with addressable network size can meet it on that company's platform rather than a competitor's.

1.1.4 *Recommender Algorithms*

Positive network externalities have a downside for users: search costs. There is a point of diminishing returns to network expansion in which those costs overwhelm the benefits of additional other users for a given user. In order to prevent this potentially growth-limiting eventuality, platforms have strong incentives to invest in tools that facilitate their users discovering whatever they happen to be looking for. This is the economic logic behind the recommender algorithms that drive phenomena like Facebook's news feed. Recommender algorithms promote both scale (by reducing the costs to users of that scale) and engagement (by positively reinforcing platform use).

Here's another way of thinking about the same problem: because users tend to be both producers and consumers of content in platforms, and successful matches are profitable (because of the skim revenue model), companies seek to provide free marketing for users in their role as producers of high-quality content, where quality is defined (problematically) in terms of the predicted desirability of that content to other users in their role as content consumers. Of course, these same algorithms are also used in paid marketing: The same identification of user interests that facilitates the placing of cat pictures on my social feeds also facilitates advertisers identifying that I am likely to be interested in buying cat products.

Recommender algorithms systems have played a key role in numerous platform governance challenges, in a variety of ways, including:

- The capacity of malicious external actors to attack those algorithms (Carley 2020, 372; Beningi, Joseph, and Carley 2019).
- The capacity of companies themselves to manipulate those algorithms for inappropriate purposes, such as associated with infamous (and highly controversial) allegations that Facebook employees de-prioritized conservative news sources in 2015 (Nunez 2016).

[11] I'm using a bunch of "may" weasel words here because I'm neither an economist nor an antitrust expert, and the nature of platform markets tends to generate debate among those who know much more economics than I do. This paragraph probably deserves an unusually large grain of salt. For a discussion of other sources of concentration from network externalities, see Martin-Laborda (2017).

- The data collection and analysis that companies must engage in to make such algorithms effective (which largely work by attempting to use machine learning to predict what users will be interested in, for various observable proxies for interest such as measures of engagement or sales), data itself which may be abused, as when Amazon has made use of third-party sales data to inform its own (competitive) product offerings (Mattioli 2020) or the infamous Cambridge Analytica scandal.
- The possibility of those algorithms to amplify pathological behavior purely as a matter of accident, as when people tend to engage with outrage-inducing content, thus fooling recommender systems into thinking they want to see more of it. For example, one leaked 2016 presentation from Facebook suggested that fully "64% of all extremist group joins are due to [Facebook's] recommendation tools" like "Groups You Should Join" and "Discover" (Horwitz and Seetharaman 2020).
- The related capacity of such algorithms to promote misinformation and polarization by meeting user demands for false or biased content. For example, one 2022 study found that YouTube's recommendation algorithm recommended more "big lie" videos about the US 2020 election to those who had already bought into Trump's election fraud narrative (Bisbee et al. 2022). As the authors suggest, this might actually be a kind of effective (albeit extremely socially harmful) design from the user and (short-term) company standpoint, since the algorithm was recommending content that those users were likely to want (Bisbee et al. 2022, 16–17).
- The danger that such algorithms may also exacerbate underlying social inequalities; for example, employment and housing recommender systems may unintentionally engage in illegal race and gender discrimination by making predictions about to whom an advertisement may be "relevant" (Ali et al. 2019).

At the same time, we can conceptualize the content moderation side of platform governance as an outgrowth of the idea of a recommender algorithm. That is, if I'm on (say) eBay or Facebook, not only do I as a user want to be able to easily find the products or posts that I'm interested in buying/seeing without having to sort through too many unwanted products or baby pictures, I also want to make sure that the products and posts that I consume are, for example, genuine rather than counterfeit, honest rather than Russian-created misinformation, and the like. In that sense, we can understand content moderation aimed at decreasing the likelihood of unwanted discoveries as no different from recommenders to increase the likelihood of wanted discoveries, in the context of limited user attention. At a sufficient level of abstraction, both recommendation algorithms and content moderation reduce to the notion of a feed ranking (as in the ordering of items on Facebook's news feed or in Google or Amazon search results) in which each item–user pair has a score predicting that user's interest and how much the platform will benefit from displaying

that item to the user, and in which users are first shown highly scored items. To moderate is to reduce that score, to recommend is to increase it.

Recommender algorithms also tend to promote a kind of winner-take-all character among users producing content or selling products: Recommender algorithms tend to reward successful attention-getting with more attention-getting.

1.1.5 *Virality (with a Caveat)*

One consequence of recommender algorithms that is distinctive to social media platforms is the phenomenon of virality, in which certain kinds of content – particularly emotionally arousing content (Berger and Milkman 2012; Brady, Gantman, and Van Bavel 2019; Guadagno et al. 2013) – tend toward extremely rapid dissemination, and hence extremely rapid and large impact. Viral content entails a number of problems. First, as an operational matter, it can outpace the efforts of content moderators to control it. Second, the phenomenon also serves as an attack surface for hostile actors, who may seek to manipulate broader publics by producing such content and seeding it widely (Bowers and Zittrain 2020, 4). Third, because such content is often of low quality (e.g., fraudulently anger or outrage-inducing), it represents a kind of overall pollution of the discursive environment of such platforms.

Although virality is characteristic of social media platforms in which content can be reshared between users – particularly at low-cost and unreflectively, doubtless drawing on what Kahneman (2013) calls "system 1" thinking – it seems to me that some variation of it is, in fact, characteristic of all recommender system-driven arrangements that rely on pre-existing user data, and thus can be characteristic of transactional platforms, hybrid platforms such as Google search, and social media without a strong resharing affordance like YouTube. Where consumption of some content or product (including, e.g., some cheap drop-shipped garbage on Amazon) is of low cost, where it is by contrast more costly to acquire information about the quality of the content or product (i.e., to exercise the psychological restraint to check on the truth of the absurd claim one is retweeting or watching on YouTube, or investigate the origins of the junk one is buying), and where a platform uses consumption as a basis for recommendation, it is possible for low-cost consumption choices to become magnified. Those low-cost consumption choices are likely to be characteristically low quality insofar as low quality is correlated with low cost, as when it is both cognitively easier and worse to consume and share emotionalizing content, or when worse products are cheaper.

I mention this because it's easy to get caught up in the micro-affordances of platforms. For example, one idea with an immediate appeal for tamping down virality is to tamp down resharing – either by removing that functionality from platforms that have it, or by impeding its use (as with Twitter's occasional interstitial that pops up on shares of news articles asking one if one wants to read the article before retweeting it). Yet in a social media world no longer dominated by chronological feeds,

intervening in user behavior is only an indirect remedy for virality, for what matters is not what users choose to reshare but what the platforms choose to display. Twitter could, for example, simply reduce the probability of unread retweets appearing on one's news feed rather than scolding its users. The discourse around virality, like that on Twitter's attempts to reduce it, seems to implicitly assume a direct causal relationship between user actions and platform outcomes, one that conceals the role of product design in that relationship.

This is an important cautionary point with respect to the notion of platform governance as a whole: The insight, associated with scholars such as Klonick and Gillespie, that governance is part of the product platforms offer needs to be matched with the equally true point that product design is part of governance. The consequences of user behavior on platforms can be altered either by interventions on that behavior or by interventions on the impact of that behavior on other users.[12]

We might go further and suggest that the notion that user behavior has a direct effect on platform outcomes is a kind of ideology, in something like a (loosely speaking) Marxist sense, that is, a false belief about the social world and its fixed and necessary characteristics which supports the interests of those who hold power and derives from those relations of power.[13] Like an ideology, it seems to pose difficulties even for otherwise sound accounts of the existing system. Just to pick a single example, almost at random, in an otherwise excellent and insightful 2018 article defending a multistakeholder conception of platform governance, Natali Helberger, Jo Pierson, and Thomas Poell (2018) state that social media platforms "only partly control content distribution" (p. 2) and "content on social platforms only spreads if many users share it" (p. 7). Both claims are false: Social media platforms completely control content distribution and could easily bring it about that content spreads without user sharing.[14]

But after abandoning the ideology that *anything* a user does on a platform has an unmediated effect on any other user, the reader may wonder why we ought to have a category of "platform governance" focused on user behavior at all. To my mind, the fundamental constraints leading to a focus on user behavior are economic: A key element of the economic value of platforms to their users is that user behavior is conferred some causal role on other users' experience – PageRank was a huge part of the value of Google's early search; the fact that one's news feed contains the cat photos one's friends have chosen to share is a huge part of the value of Facebook – and

[12] Cf. Srivastava (2021, 4) who aptly categorizes functionality like the arrangement of Facebook's news feed as a form of governance – "herding" – alongside content moderation.

[13] This is a loose version of a combination of the first and second versions of Marx's account of ideology given by György Márkus (1983). We might also think of this as a kind of reification, but Marxist social epistemology is far too deep a sea for this book to explore.

[14] For example, in 2016, the current wave of debate about alleged social media political censorship was sparked by allegations that Facebook's "trending news" feature was actively edited to suppress conservative stories (Herrman and Isaac 2016).

these economic reasons (as well as, of course, the costs of more manual curation as opposed to making use of the economic value of the "labor" of users) impose a soft constraint on the extent to which what some users are presented with on platforms and their associated external effects can simply be handled by direct intervention on platform outputs as opposed to user inputs.

1.1.6 *Diversity of People and of Behavior*

The combination of the foregoing platform features leads to a vast diversity of people and thereby of behavior. Large-scale platforms involve people with lots of different interests and resources, and hence who are predisposed to behave differently from one another. And because of the network affordances such platforms offer, many of those people find themselves interacting with others whose interests and resources are dissimilar from their own, who may have different assumed cultural starting points for interaction, and who are otherwise importantly diverse.[15] This, in turn, leads both to behavioral novelty when those diverse interactors discover shared behaviors compatible with their interests, resources, cultural backgrounds, and so forth, and to conflict, when they do not.

This is a kind of converse of the niche-finding property, and one of the most interesting features of platforms is that they can go together. However, this is not wholly novel – it's also true of, for example, the city – in which the sheer scale and density of the urban environment mean that one can both find one's niche interests (because there are enough people to economically support serving them in a big city) *and* that when one is out and about on the street one is liable to be exposed to people who are different from oneself, and perhaps learn from one another and invent new ways of being in the world together.[16]

We should distinguish behavioral diversity and user diversity. User diversity is a consequence of scale. Behavioral diversity, however, is both a consequence of user diversity in the context of recommender algorithms promoting niche-finding and the simultaneous fortuitous interactions between different users who would never interact in other contexts. But they're also a feature of certain kinds of platform designs, themselves a consequence of the incentive to promote engagement.

Thus, the interactive affordances of platforms tend to be open-ended within the boundaries of the technical forms available to a company: YouTube doesn't exist to distribute specific kinds of videos; leaving aside concrete law or policy-based restrictions, it exists to distribute whatever videos its users see fit; the same can be said about eBay and goods.[17] The economic incentives for scale and engagement suggest

[15] Moreover, users seeking a market for their unique offerings (whether content or products) are, assuming an effective recommender algorithm, much more likely to find that in a larger network – hence there is probably an incentive for users with niche offerings to flock to larger networks.

[16] On the relationship between density and diversity, see, for example, Moroni (2016).

[17] Sometimes this is called "generativity" (Gorwa 2019b, 3).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

that a high degree of adaptivity to diverse user needs will be rewarded – and this further suggests that one key economic incentive of platforms will be to diversify the technical forms available in order to make additional types of interactions available to their users.[18]

The foregoing implications taken together lead to a high likelihood of interactions on multilateral platforms that tend to be unanticipated and unanticipable in advance. Content goes unpredictably viral. Emergent properties of densely networked interaction at scale crop up, seemingly out of nowhere. The broader ecosystem of content on such networks can be subject to high-speed evolutionary landscapes where particular kinds of activity propagate and disappear. The phenomenon of "context collapse" upsets people's ordinary intuitions about the audiences for their activity and hence leads to surprise interpersonal responses even on an individual basis (Marwick and boyd 2011) and then that scales too, and you end up with, well, Twitter.

1.2 PLATFORM PROBLEMS ARE LIKE STATE PROBLEMS: THERE ARE LESSONS TO BE BORROWED

In consequence of the above, platforms tend to pose severe challenges for the control of malicious or unintentionally harmful behavior. Novel kinds of behavior emerging at high speed among a vast and diverse populace in the context of positive feedback loops on attention make it difficult to identify the harmful behavior likely to appear, and difficult to enforce rules against identified types of harmful behavior.[19]

Those properties, in other words, generate distinctive governance difficulties that have been observed worldwide in recent years on social media. For example, the phenomenon of viral hoaxes has been particularly hard for such platforms to control not merely because there are internal and external conflicts about the role of such platforms in policing truth (although this is certainly the case), but also because such hoaxes can exist in a huge number of languages, in swiftly evolving idiomatic form, distributed by a huge number of users in a huge number of ways – and thus

[18] In particular, this incentive seems to drive at least some part of both Google's and Facebook's product strategies, such as with Google's persistent experimental creation (and routine destruction) of new ancillary services, like Google Reader and Google Flights, as well as acquisition of companies occupying different roles in interactive contexts, like Blogger, as well as Meta's acquisitions of companies specializing in different technical forms, such as virtual reality, and the expansion into new forms of interaction such as trade ("Facebook Marketplace") and dating. Schwarz (2017, 384) plausibly suggests that this tendency also arises from the monetization of user data, as it creates new profit opportunities derived from merging data across novel domains of user activity.

[19] Platforms attempt to leverage the same artificial intelligence techniques driving their recommender systems to also identify problematic behavior at scale and speed. See, for example, Mark Zuckerberg testimony before Senate Commerce and Judiciary Committees, April 10, 2018, available at www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/ ("Now, increasingly, we're developing A.I. tools that can identify certain classes of bad activity proactively and flag it for our team at Facebook").

because any platform wishing to control those hoaxes would require the capacity to identify their distribution in these ever-changing forms at speed and scale – a gigantic problem. Moreover, those platforms must be capable of developing policies that describe forbidden behavior at a sufficient level of generality to be administrable without making countless costly and inconsistent one-off discretionary decisions, but which are still capable of being sufficiently tailored to individual situations to achieve the underlying goals of those policies.

In these respects, platforms bear an important functional isomorphism to key forms of political organization such as cities and states. Polities are also, on some important ways of thinking about the role of political organization, fundamentally infrastructural – polities provide a broad array of affordances, such as physical security and enforceable contracts, in order to permit people to engage in diverse behavior at scale. This is a notion that's shot through legal and political theory, or at least the small-l liberal kind – if there is a foundational proposition of liberalism, it is that the function and justification of the state is to help people to engage in private projects. For example, some scholars think that a key function of law is to solve coordination problems getting in the way of people achieving their collective well-being – to do stuff like enable people to decide which side of the road we all drive on (e.g., Fuller 1969; Postema 1982, 182–86). A related way of thinking about the function of law and of the state is "determination," identifying principles of (property, etc.) law to be enforced in order that people may have secure access to means by which to pursue their private ends (e.g., Ripstein 2004, 11–13, 26–29; Beitz 2018, 425–27). Contract law is often said to be explained and justified in view of the fact that it permits people to make complex, binding, plans relying on one another, and thus solve all kinds of problems of trust and credible commitment that would otherwise hinder socially and individually useful joint projects among individuals (e.g., Dagan and Heller 2020). Political philosophers can recognize this idea in Rawls's supposition that the original position models a decision-making process motivated by the goal of framing a society in which its members can develop and pursue their conceptions of the good – and that the products of that society are divided as the "benefits … of social cooperation" (Rawls 1999, 4).

In that sense, one could argue that *the state itself is a platform*.[20] Moreover, the interests of states align with the interests of their citizens in the same broad way that the interests of platforms align with the interests of their users. States typically derive

[20] Some in the technology industry have offered a version of this suggestion. For example, software industry publisher Tim O'Reilly (2011) analogized government to hardware platforms insofar as they offer infrastructure for collective action among their people. Arguably, instead of the state, we might conceptualize the law itself, in contexts including, but not limited to, states (e.g., Ellickson 1991) as a platform, though I cannot explore this idea here. It is interesting to note that law has positive network externalities to the extent that the transaction costs of far-flung and diverse economic and social actors are reduced by having a common law to which they may appeal – a dynamic that contributes to some economic historians' accounts of the *lex mercatoria* in medieval Europe (Milgrom, North, and Weingast 1990; but see Kadens (2012, 2015) for historical dispute as to its existence).

revenue from taxing productive activity; one influential theory of how states learn to refrain from recklessly expropriating and abusing their citizens observes that to the extent rulers can extract rents from that activity, they do better to provide their people a minimum of security in order to facilitate the activity's occurrence (Olson 1993).[21] At the highest level of abstraction, advertising revenues on Meta or commissions on eBay are a kind of a tax on user activity.

We might further say that just as moderation is part of the product of companies, the rules of property and contract and the like are part of (or the fundamental) product of governments. In fact, there have been cases where governments or rulers have directly sold access to the legal form of resolving disputes. For example, medieval English kings received meaningful revenue from the sale of adjudication, either in the form of court fees or in the form of kickbacks from adjudicating cases promptly (Ramsay 1925, 1:1–2, 58; MacPhail 1926, 251). In the contemporary world, we can think of paid private arbitrators as the market supply of adjudication undersupplied by the state. Contemporary court fees also represent, in some sense, the idea of legal resolution having a monetary value that the state can partially exploit.[22] Moreover, contemporary choice of law and choice of forum rules and their use in the commercial world might represent the idea that states compete in their supply of legal rules and adjudication.[23]

Similar points are true of sub-state forms of political organization. For example, a city, understood as a collection of affordances for social interaction analyzed by influential urban theorists spanning a range from Jane Jacobs to Richard Florida, resembles one of the multilateral platforms described above. The success of a city is a consequence of its ability to promote beneficial social and economic activity by bringing together, at scale, individuals seeking to participate in that activity with one another, and cities provide affordances in the form of physical spaces such as roads and parks, as well as incentives, such as economic development grants, in order to make these interactions possible; cities also fund their activities by taxes qua skim off the top of such activity. The density (which can be represented as the potential for each resident to have more network connections with others due to sheer physical proximity) of a city itself provides some of its key advantages, both in terms of practical economies of scale and in terms of positive network externalities from, for example, locating sellers and buyers and workers and jobs in the same place. Cities also tend to promote diversity and novel forms of behavior and, with that, novel

[21] Contemporary officials in modern democratic states, who do not personally receive a share of taxes on productive activity, still have an incentive to facilitate that activity (a.k.a. "economic growth") to the extent that electoral advantages to prosperity substitute for the direct receipt of tax revenues by individual leaders.

[22] Some American municipalities use court fees charged to those who encounter the criminal justice system as a source of revenue (Sances and You 2017; Mughan 2021). The US federal courts also have sold access to court records at a price which critics have alleged exceeds the cost of operating the system (DeWitt 2010).

[23] Though this competition is sometimes, perhaps even usually, pernicious (e.g., Slemrod and Wilson 2009; Anderson 2015).

enforcement and safety challenges – plus the same density that helps sellers find buyers also helps criminals find victims.[24]

Unsurprisingly, given that they have (at a certain level of abstraction) similar incentives and serve similar activity-enabling functions in the lives of their people, polities have experienced many of the same governance challenges with which platforms now struggle. For example, perhaps the most fundamental challenge of using the form of law to govern human behavior is to strike the balance between general rules to be consistently enforced and the capacity to adapt to specific situations in the universe of potentially infinite human behavioral novelty (e.g., Schauer 1991). This is a problem for polities and platforms alike. In the abstract, the balance between general rules and adaptability to specific situations in application is a problem associated with *all kinds of behavioral regulation* – in principle, for example, it applies to situations as prosaic as regulation by parents or employers (how to set attendance rules at work that are both fair to all, yet accommodate individuals' special circumstances?).

Theorists in the domain of polities have wrestled with these problems for a long time. The management of open-ended interaction affordances in the context of scale and diversity has been a central concern for political theory as long as that discipline has existed. In Ancient Greece, for example, Aristotle famously wrestled with the problem of scale, suggesting that a polity ought to be small enough that its members know one another's characters, in order that they might effectively select officials and conduct litigation.[25] Contemporary scholars have identified that scale and diversity entail distinctive problems of central state planning in that central policymakers will often lack knowledge to be found only on the edges of the community; important academic work focusing on this problem has ranged from identifying it as the source of disaster in modern governance (J. C. Scott 2008) to favoring markets as a solution to information problems (Hayek 1945), to explaining the success of democratic institutions for aggregating dispersed information (Ober 2008) (much more on all of these in Chapter 3).

[24] Physical economic markets also have something of this platform character. Economic historians have written about institutions such as the Champagne Trade Fair, a physical site (or set of them) – supported by law but having a physical importance in its capacity to bring trading partners together. Consider also the New York Stock exchange and more broadly the contemporary electronic stock trading system, qua infrastructure for making trades, even novel trades that can be adapted to functional forms afforded by the system. Such markets again seem to partake of positive returns to scale, with returns that are internalized by market operators and those that are bestowed on participants. Contemporary financial markets also seem to partake of phenomena that bear at least a superficial resemblance to virality. For example, it may be sensible (I am not sure) to represent feedback effects associated with "flash crashes" as a kind of viral propagation of toxic false information about some market participants, represented in the prices associated with such participants (on flash crashes, see, generally, Lange, Lenglet, and Seyfert 2016). Of course, this recently merged with the problem of social media as Reddit created a kind of "flash bubble" in stocks like GameStop. Such markets also produce beneficial as well as pathological novelty, including, for example, the exotic financial instruments that were used to conceal the riskiness of subprime mortgage loans in the mid-aughts. Maybe some of the policy proposals described in this book could also apply to stock exchanges, though I don't dare try to defend such a notion here.

[25] Aristotle, *Politics*, book VII, 1325b–1326b.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

The benefits and challenges of diversity at scale are an axis around which debate has revolved in areas as widespread as the provision of social welfare benefits (Kymlicka 2015), the creation of patriotic identities (Müller 2007b), and novelty-driven social progress and economic growth (Thomas and Darnton 2006). One fair description of the first premise of contemporary liberal political philosophy is that its goal is to describe political institutions that permit a large population that is diverse in both interests and values to live together while maximizing the freedom of each individual to pursue their individual interests and values. This is, for example, John Rawls's (1985, 233) conception of the "two moral powers" of each citizen. Almost every public good provided by a political institution can be described as a set of affordances to pursue the development and exercise of those moral powers. On the opposite side of modern liberal theory, Robert Nozick (1974) implicitly applied the idea of positive network externalities and returns to scale in his story of how a libertarian state might arise from a "dominant protective association."[26]

Given the broad array of functional similarities between internet platforms and polities, it should be unsurprising that ideas from liberal political philosophy have leaked into the self-understandings of platform users as well as the companies themselves. The most salient example is that of the idea of free speech – there is persistent controversy in the United States about the extent to which First Amendment rights ought to apply against decisions of those platforms.[27] At various points in their corporate histories, social media companies have proudly trumpeted their free speech credentials (Gillespie 2010, 356), with Twitter famously describing itself as "the free speech wing of the free speech party" (Halliday 2012); and positive accounts about the impact of social media have focused (at least prior to 2016) on its capacity to facilitate core First Amendment-type values such as politically dissident speech in autocracies (e.g., Passini 2012).

Nor are those normative claims necessarily as misguided as they might seem from the perspective of a frustrated 2023 civil rights lawyer sick of explaining to people why there's no First Amendment right to spout racial slurs on Facebook.

[26] Early in the internet age, David Post (1995) borrowed from Nozick's idea of competition between providers of rules to sketch out a vision for the development of governance of online activity.

[27] The simplest answer in terms of existing legal doctrine (which is more of an "is" than an "ought" issue) is "not at all," although there is an important, although largely disused, corner of First Amendment doctrine according to which the operators of the "functional equivalent of the traditional public square," such as owners of company towns, may be treated as quasi-governmental actors subject to the First Amendment. See, for example, *Marsh v. Alabama*, 326 U.S. 501 (1946); *Robins v. Pruneyard Shopping Center*, 23 Cal.3d 899 (1979). Scholars have analogized social media both to cities in the form of the town square and to governments in the form of speech regulators in order to consider the possibility that such doctrines may be applied to Facebook and Twitter (e.g., Tussey 2014). There has also been legislation reflecting this position passed in both Texas and Florida, and as of this writing there are ongoing controversies about the constitutionality of those statutes (which controversies are, blessedly, beyond the scope of this book).

Scholars have recognized that "the state" is not an organic entity that can unproblematically be defined from outside – as Bevir and Rhodes (2015) explain, the state is a kind of sociological and narrative phenomenon (in their words, "cultural practice") derived from the self-understandings of participants in activities of governing. Similarly, international relations scholar Swati Srivastava (2022) has defended a theory of "hybrid sovereignty" according to which private organizations ranging from the English East India Company and Blackwater's mercenaries to nonprofits like Amnesty International have become entangled with allegedly core aspects of state sovereignty or even have such aspects delegated to them.[28] Blackwater, for example, carried out combat in Iraq on behalf of the US government, defying the notion that the use of military force is the absolute heart of modern indivisible state sovereignty. Platforms do something similar. The intermediary liability framework of statutes like the DMCA, for the most obvious example, seems to recruit platforms to carry out copyright adjudication and enforcement on behalf of the state in a manner not terribly dissimilar from how Blackwater was recruited to carry out war, except that the tool used to do so was the risk of liability rather than contract (i.e., a stick rather than a carrot).[29] Human rights scholar Molly Land (2019, 404–8) has already argued that these intermediary liability frameworks are a form of delegated state authority for some international law purposes.

In my view, there's no a priori reason to rule out platform companies from being part of such a state practice or hybrid sovereignty, particularly when their personnel are using familiar forms of law-like reasoning to issue authoritative determinations about individual behavior in their limited contexts – and hence no a priori reason to deny that platforms are subject to the normative standards we apply to states, although our judgment about those normative standards should not be taken to determine the separate (though obviously entangled) question as to whether legal standards such as the US First Amendment should apply to them. This book ultimately takes no position on whether platforms ought to be treated as carrying out some form of state action or manifesting hybrid sovereignty, which is not a problem that needs to be resolved for the argument to go through. But I *do* contend that it should be clear, perhaps even uncontestable, that platforms do carry out some tasks that are meaningfully similar, in both their social role (e.g., managing political discourse) and in the constraints they face (i.e., managing scale and diversity), to tasks carried out by states.

[28] For a much more sinister example, I have elsewhere suggested that white supremacist paramilitary organizations had, in some legally and theoretically relevant senses, become part of or taken over part of the state in at least some time periods in the US South after the end of Reconstruction (Gowder 2021, 92–93; 2018a, 348–51).

[29] It may be that intermediary liability can serve as a gloss on the form of hybrid sovereignty which Srivastava calls "contractual hybridity." It should perhaps be unsurprising that Srivastava (2021) has also written one of the few articles from the institutional side of political science offering serious attention to the notion of platform governance as governance.

Those observations suggest that efforts, whether governmental or private, to build institutions capable of shaping aggregate behavioral outcomes on networked internet platforms can benefit from drawing on our existing knowledge (theoretical and practical) of political and legal institutions.[30] Indeed, such efforts are already ongoing – it is not a coincidence that organizations such as Facebook and Google, for example, recruit lawyers – members of the profession built to handle precisely such governance problems – to head their content policy teams,[31] or develop organizational forms explicitly derived from governments.[32] Nor is it coincidental that scholars such as Kate Klonick (2018, 1641–48) have explicitly analogized these efforts to the acts of government policymakers and identified that the process of actually implementing platform rules looks remarkably like administering a legal system. It so happens that the functional forms of legal systems developed as responses to problems remarkably like those which platforms face.

For those reasons, it should now be clearer why I contend that theoretical approaches to political and legal institutional analysis may help us understand the problems of networked internet platforms. Moreover, I shall suggest that the institutional tools of polities, such as the rule of law and techniques associated with ecological governance and urban design, may help us alleviate them.

[30] Cf. Gorwa (2019a, 859–60), observing a shortage of input from political science in platform governance.

[31] For example, former federal prosecutor Monika Bickert leads a team making content policy at Facebook (Zuylen-Wood 2019) (for purposes of full disclosure: Bickert attended Harvard Law School at the same time as I did, though we haven't spoken in many years). Google's head of "Trust and Safety" is not a lawyer, but is a White House veteran (Canegallo 2019). Other high-ranking personnel in Google's/Alphabet's content policy operation have been lawyers; for example, a former ACLU lawyer occupied a number of leadership roles in YouTube content policy. (See the LinkedIn Profile of Juniper Downs, www.linkedin.com/in/juniper-downs-5245879/, visited April 20, 2020.) According to a profile by the India Times, Twitter's pre-Musk general counsel bore responsibility for its content policy (India Times 2020).

[32] For example, Meta's content moderation oversight board was conceived as a kind of "Facebook Supreme Court" expressed in exactly those terms (Ingram 2019).

2

The Enterprise of Platform Governance Development

Transnational organizations and hegemonic donor states such as the United States have concerned themselves in recent years with building the state capacity of developing and transitional states. The United States made efforts to develop the military and police capacity of postwar Iraq (after tearing down its prior version of those institutions) (Gates 2010). The European Union has made efforts in Eastern Europe to build anti-corruption capacity (Börzel and van Hüllen 2014). There have been substantial global efforts to build state capacity in Somalia (Menkhaus 2014). At least to some extent, these efforts are motivated by the self-interest of those leading the capacity development efforts – the more functional Somalia's state is, the more it can control the piracy that harms its neighbors. And it may be that a functional Somali state would be better at controlling piracy than external states trying to do so themselves, for example, because of its superior local knowledge or sociological legitimacy. In other words, the enterprise of capacity development is often predicated on the ideas that governments that are unable or unwilling to control their own citizens create *negative externalities* for the rest of the world, and that the rest of the world is unable to control those behaviors directly, without helping the state with direct access to the people causing the problem to build the capacity to intervene.

This chapter argues that, from the standpoint of government regulators, analogous incentives exist with respect to platforms. Governments have strong reasons to control much of the pathological behavior that appears on networked internet platforms, from commercial fraud on Amazon to fake news on Facebook (although the example of fake news also illustrates that government reasons are far from unequivocal and may be outweighed by other interests, such as individual freedom). And it is likely that governments may be ineffective at directly regulating some of this behavior, for example, because of the tendency of platforms and their users to span jurisdictional boundaries, because of the property rights of owners of those platforms, or simply because of the bureaucratic burden of regulation without direct access to the immediate site of behavior in the form of platform code and interactional surfaces. Such governments accordingly have reasons to promote the

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

development of a kind of analogy to state capacity in platforms in order that they may require or encourage those platforms to control behavior that appears via their infrastructure.

2.1 WHY IS CONDUCT ON PLATFORMS SO HARD TO REGULATE?

Some of the ways in which many commentators worry about externalities caused by platforms represent what we might think about as pure conflicts of interest between those platforms and society at large. Social media companies tend to profit by collecting large amounts of information about user activity and using that information for ad targeting. That is (arguably) bad for society as a whole, insofar as it allows advertisers to manipulate consumers and leads to the risk of dangerous third-party data leaks. But it's not bad for companies (the data leaks are – but the manipulation is precisely the service they offer to advertisers). Similarly, "sharing economy" transactional platforms tend to make money by evading regulations governing services like taxicabs and hotels as well as workplace law in general. Again, that's a problem for the rest of us. Under such circumstances, the best we can do as a society is use the legal system to simply coerce companies to act in ways that are less self-serving and more public-serving by, for example, requiring them to count their drivers and delivery people as employees rather than contractors or prohibiting them from sharing user data between different corporate functions.

By contrast, other kinds of external harms – the sorts within the domain of interest alignment described in the Introduction – likely also harm platform company interests, at least under certain interpretations of those interests. The clearest examples involve various kinds of fraudulent behavior which can harm companies' long-run interests in running platforms that their users can trust for reliable information. For example, viral disinformation on social media undermines the implicit value, at least over the long run, that those users can attribute to information found on those platforms. The same goes for counterfeit products and fraudulent ratings on transactional platforms.[1]

With respect to such harms, it is a genuine puzzle (which this book aims to address) as to why the companies cannot simply stop them. This puzzle is highlighted by the

[1] To the reader who is skeptical about the claim that company and social interests are aligned in these cases: Wait a few paragraphs. I absolutely acknowledge that sometimes, company interests are more ambiguous or conflicted than may appear on the surface, and Chapter 4 addresses such cases at length. It may be that the harmful behavior under examination, for example, promotes short-term profit (by generating social media platform engagement or transactional platform sales) at the expense of long-term profit; that different constituencies within a company's employees and/or ownership differ as to their evaluation of the relative benefits and burdens of such activity; or that companies operate under external pressure (such as by government officials who benefit from social media disinformation) which effectively changes platform interests by giving them an incentive to permit the harmful behavior. But sometimes, as in the Myanmar case which opened this book, company interests are not conflicted, yet they still cannot stop the behavior.

fact that there are important senses in which technology platforms have vast governance advantages over physical world governments.

First, unlike physical world governments, company personnel can observe anything (though not everything at once) that happens under their control (with the sole exception of end-to-end encrypted platforms like WhatsApp – but even there, the encryption is itself a platform design choice under company control). They need not install surveillance cameras and peek through windows, they simply need to query a database. Such observation is not costless – it still requires paying personnel to actually examine some piece of on-platform behavior (a product for sale, a tweet), but the costs are substantially lower per observation than for states to send police to follow people around, get search warrants, and so forth. Moreover, some (but not all) "observations" in the platform context can be automated by, for example, deploying artificial intelligence to identify certain kinds of conduct. While artificial intelligence and other automated methods of observation are no panacea, they have certainly proven effective in a number of important cases. For example, companies have devised methods to automatically fingerprint and then identify duplicates of known child pornography images so that they may be removed from platforms without any human having to look at them.[2] It is also possible to have a kind of human-in-the-loop observation where artificial intelligence identifies behavior to be reviewed by a human. Governments can automate enforcement too (speed cameras exist), but governments still have to expend costs to translate behavior in the physical world into data, whereas companies have data all the way down.

Similarly, many forms of rule enforcement are, at least in the abstract, much cheaper for platform companies than for physical world governments. The most common tools of enforcement in the platform context are various forms of restriction or removal either of user accounts or of the specific conduct of user accounts (such as posted content). But, once again, these enforcement techniques do not require the building of prisons or the operation of courts, or the sending of heavily armed police officers to people's houses – they simply require the design of software to create affordances for enforcers within companies to alter user privileges and content, and then the operation of those tools – the click of a button within some internal system to change a field in a database.

But the relative ease of governance of platforms is to some extent an illusion, for platform companies are still subject to two core abstract-level problems that they share with physical governance.

The first is the difference between observation and knowledge, namely interpretation or what I will describe in Chapter 3 as "legibility": While platform companies can "see" any behavior on the platform by querying a database, interpreting that behavior, and in particular, determining whether that behavior is harmful or

[2] See Microsoft's description of the PhotoDNA technology at www.microsoft.com/en-us/photodna.

harmless, compliant with company rules or noncompliant, is much more difficult. In many ways, this interpretation is likely to be more difficult for platform companies than physical world governments, for platform companies – particularly the most successful companies at the greatest scale – have to deal with a vast amount of behavioral diversity and novelty, and are much more socially and culturally distant from many of their users. A bunch of lawyers and engineers in Menlo Park are likely to have a much harder time understanding the behavior of people in, say, Luxembourg or Kenya than the local government of a municipality composed of representatives of the people they govern.

The second is the problem of incentives. Not every person within a company's decision-making structure may have the incentives necessary to actually govern problematic behavior. Personnel whose job is to keep government officials happy may be more vulnerable to external pressure by politicians who benefit from disinformation. Engineers and product designers on "growth" teams who are rewarded for increasing activity may have an incentive to build features that facilitate all activity, whether those are harmful or helpful to a company's long-term interests. Stock markets may give executives an incentive to permit behavior that promotes short-term profits at the expense of long-term profits.

Those two problems are, respectively, the subjects of Chapters 3 and 4 of this book – I contend that they are the core challenges that vex the enterprise of platform governance in the domain where company interests (properly understood, from a long-term perspective) and social interests are aligned.

2.2 WHY BUILD CAPACITY?

The global community has long wrestled with the problem of states that are unable or unwilling to regulate the harms that users of their territory inflict on the rest of the world. Piracy is a classic example: When a state fails to adequately patrol its territorial waters, it can become a haven for pirates who then prey on international shipping (e.g., Pham 2010). But there are other kinds of criminal activity that can result from governance failure (or governance unwillingness) and can predominantly affect people outside a given country, ranging from terrorism to intellectual property misuse and the cottage industry of internet scams in certain towns in Eastern Europe (Subramanian 2017; Bhattacharjee 2011).

This problem may arise because a country lacks sufficient ability to fully regulate behavior on its territory. Maybe, for example, the government lacks legitimacy with the population, or the funding to train and deploy adequate security forces. Moreover, it may lack adequate incentive, either simply in virtue of the fact that much of the cost of the harm is externalized or because leaders have mixed motives, as when the criminals pay bribes to continue their activity. The piracy example is again apt: Percy and Shortland (2013) have plausibly argued that even improvements in state capacity are ineffective in controlling some cases of modern

piracy in view of the fact that the activity provides local economic benefits and its harms are externalized.[3]

Nonetheless, it may be difficult for other states to directly regulate this behavior, even though it inflicts harm on their citizens or in their territories. They may not, for the simplest example of a reason, have their own physical access to the territory where the international harms are being generated. And they may lack sociological legitimacy with the people of that territory, even more so than an unpopular domestic government – it should be unsurprising, for example, that the people of Somalia would be unwilling to be governed by the United States and France.

Accordingly, a core strategy that other states have adopted to address the problem of externalities generated by under-governance is to attempt to assist the state which is failing to govern in building or rebuilding its capacity to do so, and providing it with incentives to use that capacity. On our running example, Bueger et al. (2020) describe maritime capacity-building efforts in various nations in the region to address piracy from Somalia. Similar efforts have been made in order to improve the ability of countries to fight international terrorism within their territories (e.g., Bachmann and Hönke 2010).

Observe the incentive structure underneath such capacity-building enterprises. Under-governing states create certain kinds of external harms in view of the fact that they fail to control the behavior of people operating from their territory. For convenience, I label those kinds of harms *governance failure externalities*. Moreover, such states not only fail to control harmful behavior, there's an important sense in which they provide resources with which individuals may carry out that behavior, and hence provide an incentive for that behavior. The most obvious example of such resources is a headquarters – by allowing (deliberately or inadvertently) pirates or terrorists or international fraudsters or whomever to operate from their territory, such states shield them from the direct control of others – Westphalian notions of sovereignty impose a barrier to more effective states directly acting to control the harmful behavior. Another related example of a resource that failed states provide is access to victims – by failing to control an important region of the ocean located among an important trade route, a state effectively provides pirates with a known location in which victims may be found. Regardless, the point is the externalities: other states, and international organizations, have self-interested reasons to provide capacity-building assistance and incentives to lower the cost of the home state regulating domestic behavior with negative spillover effects.

[3] A more sophisticated description of the problem that captures this point may be that strong states which do not have a problem with piracy have little incentive to permit it, as they can prosper from less harmful forms of economic activity; however, once pirates get a foothold in a weak state, it may be that when such states become stronger, they still lack adequate incentive to expend the costs to shift to more nonviolent and less harmful forms of revenue generation.

I do not wish to describe this problem in a way that buys into problematic US-centric or developed country-centric representations of global problems and responsibility. Wealthier and more powerful countries can also impose governance failure externalities on the rest of the world. For the most obvious example, the failure of the United States to control the carbon emissions emanating from its territory imposes vast harms on the rest of the world. In terms of long-run global welfare, greenhouse gases are far more damaging than pirates. In principle, it would make sense for the rest of the world to help us Americans control our polluters. Unfortunately, given the vast power and wealth disparity between the United States and most other nations, our troubling belief in American exceptionalism, and the fact that our governance failures are built deep into our polarized politics and ineffective constitutional institutions, there's probably precious little that other countries could do to help us.[4]

Platforms create precisely the same kinds of governance failure externalities. Here, some care is in order: a corollary of the point which I have repeatedly emphasized that only some platform harms are within the domain of interest alignment with society is that not every harm that platforms impose on the outside world is a governance failure externality. Sometimes, the platforms themselves do things that directly harm third parties – such as when they violate people's privacy or squash their competitors. But other times, the externalities come from the independent behavior of users making use of platform affordances.[5] In the platform context, the most obvious examples revolve around various kinds of fraud or confusion – when viral disinformation circulates on social media, for example, or when counterfeit N95s plague Amazon. And much like in the international context, the affordances of platforms both facilitate this behavior and make it harder for governments to control (primarily by increasing its quantity and speed, as well as cross-border impact).

[4] Moreover, the politics of the United States are also creating externalities even in the domain of social media: terroristic conspiracy group QAnon, which arose in the United States (perhaps with Russian assistance) and spread in social media on behalf of Donald Trump (Bleakley 2021), has spread to other countries, notably Germany (Hoseini et al. 2021), where it was linked to a coup plot in late 2022 (Gilbert 2022).

[5] As noted in the virality section of Chapter 1, there's an important sense in which *everything* platforms do comes from the platforms, insofar as they control the causal mechanisms that permit third-party behavior to have an impact on others, such as the distribution of content over social media. But, taking as given an economy of platforms in which user behavior is conferred some (platform-mediated) causal impact on the experience of others, we can coherently talk about categories of harms where it makes sense to attribute primary responsibility for those harms to users. For example, suppose that some method of content distribution has many socially valuable uses, such that it would sacrifice significant public well-being to ask a social media platform to disable it, but that method of content distribution also has some small number of seriously harmful uses. Under such circumstances, it seems to me most plausible to point to interventions on user behavior to end the harmful uses as the best path forward for overall social well-being, and hence to attribute responsibility for harms generated in the absence of such interventions primarily to user behavior.

More specifically, the relative inability of governance "donor" states to act directly on the people of recipient states is replicated in the platform context. As a whole, companies tend to have both access to more information about user behaviors and access to a wider array of tools (such as major interventions on their own platform design) to regulate them than do governments. By contrast, even though ordinary platform users would presumably prefer to be able to use their market power to mitigate some of the harms they receive from platform companies (especially if that sometimes means, e.g., accidentally buying fake N95s in a pandemic, or being deceived by political lies), governments have much more leverage over platform companies than the users of those platforms do, in virtue of the fact that governments, of course, have more coercive tools available (they can freeze assets and lock up executives). Users have very little power over platforms themselves (their only viable option is exit, network externalities lead to very high switching costs for individuals, and mass coordination is costly).

This suggests that a regulatory capacity-building approach dedicated to the goal of empowering companies to protect the interests of users and society – which this book argues will also require empowering ordinary people to intervene on platform choices – is an appropriate way to think about how we, as citizens of democracies wishing to control our environments, ought to bring ourselves, as users of these platforms, under control. Of course, there are dangers as well to the project of capacity building – which I will address below – most obviously the rightful fears that many of us have about enhancing the power of unaccountable private platform operators over behavior in a context where increasing shares of important human interaction happen via those platforms. But only when we enable ourselves to think directly about governance capacity do we even become capable of properly raising those questions.

2.3 HOW TO BUILD CAPACITY?

In the abstract, it is helpful to think about the enterprise of capacity building as giving a putative governor (as a first pass, a platform company – but the remainder of the book will call into question the notion that it is *companies* in particular who should be doing the governing, as opposed to a combination of companies, workers within companies, civil society, perhaps governments, and most importantly ordinary people) both the *incentive* and the *tools* to control unwanted behavior.

In the context of states, relevant examples of such capacity-building efforts include the enterprises of rule-of-law development and of security assistance. Powerful states and their associated NGOs have offered less powerful states a variety of tools for more effectively extending their authority over their territories, including military training and arms, legal personnel training, and building of physical infrastructure for legal administration such as courthouses; of particular relevance for the argument to come, some of these resources (such as training) have also been extended

to civil society actors within recipient countries rather than governments directly, with some success in achieving the goals of legal development (Blattman, Hartman, and Blair 2014). Rule-of-law development efforts also have intervened directly on state incentives, such as with constitution-drafting projects; additionally (and rather more controversially) economic development programs directed by entities such as the World Bank and the International Monetary Fund have also aimed to intervene on state incentives with "structural adjustment" programs which impose obligations on states to do things like privatize public services as a condition for receiving international loans.

Governments have analogous tools with respect to companies. The United States or Europe could, for example, impose changes on the corporate structure of platform companies or on the ways that decisions are made within that structure by alterations to corporate law. In other corporate contexts, we have seen examples such as the Sarbanes-Oxley Act, enacted after the Enron scandal, which imposed new requirements on the boards of directors and outside auditors of public companies. We can understand this as directly analogous to (but hopefully more coercive than) the constitution drafting assistance offered by the world's liberal democracies in places such as Afghanistan. Foreshadowing the proposals articulated in Chapter 6 as well as this book's Conclusion, governments could also directly insert third parties into the governing structures of companies, for example, by mandating certain kinds of representation on corporate boards, by giving third parties a veto on some company decisions, or even by directly acquiring company stock and exercising decision rights on behalf of empowered third parties. This too is a governance tool with which governments are familiar. For example, Article 27 of the European Charter of Fundamental Rights provides that workers have a right to "information and consultation." As implemented by an EU directive, workers are entitled to that "information and consultation" on, among other things, "decisions likely to lead to substantial changes in work organisation."[6] Member states have further implemented this right by providing for worker consultative bodies within companies.[7]

While there are limits, such as the American constitutional protections for property (and, in the case of social media companies, free speech), the very brief survey above should suffice to illustrate that governments have the tools to intervene quite deeply into the decision-making processes of platform companies. In view of the significant (and legitimate) coercive power that countries exercise over those

[6] Directive 2002/14/EC of the European Parliament and of the Council of 11 March 2002 establishing a general framework for informing and consulting employees in the European Community – Joint declaration of the European Parliament, the Council and the Commission on employee representation, Article 4(2)(c).

[7] See, for example, French Code du travail, Articles L2311-1 to L2311-2, which sets up a social and economic committee ("comité social et économique") in companies above a certain size to represent employees; that committee has a broad array of powers including receiving substantial amounts of information from the company, raising complaints to board level, investigating safety issues, and advancing employee concerns to government officials.

companies, we also have reason to think that the use of those powers for purposes of capacity development has greater prospects of success than capacity-building projects in the context of states.

This book concerns itself with the enterprise of governance in the abstract, that is, with the kinds of institutions (organizational forms) that ought to be adopted in order to permit platforms to effectively govern (and permit us to effectively govern them). In the context of the analogy to governance development in the international community, this is analogous to saying that the scope of concern should be the promotion of institutions like democracy and the rule of law, rather than the prescription of the particular substantive legal and economic rules which recipient countries ought to adopt.

The institutional approach recognizes that the concrete rules that platforms – or even platforms in specific sectors, such as social media – ought to adopt will potentially vary widely across platforms, across underlying social contexts, and across time (just as with countries). Moreover, many of the real-world rules problems facing platforms are sufficiently contested and potentially irresolvably difficult in their form as rules, even if there are correct answers to them. What, for example, is the normatively correct position for a platform to take on disputed claims about treatments for COVID-19, or allegations about the behavior of a US president and their family, for example? It seems to me overwhelmingly likely that the best answer to this category of problem will operate in a combination of the domains of what Rawls (1999, 73–78) called "imperfect procedural justice" and "pure procedural justice." That is, that the answer to the question will focus on how the decision is to be made (and, by extension, who is to make it), rather than what the decision is to be, sometimes because the procedure is our best albeit imperfect way to get to an externally defined right answer, and sometimes because whatever answer follows from the procedure (as with democratic policy choices) is for that reason the right one.

Another component of the institutional approach adopted in this book is the supposition that the concrete interventions that are likely to be effective in platform governance are more in the domain of product design in terms of the high-level affordances available to users as well as in tools and methods of governance, rather than in low-level enforcement details or volumes of work carried out (such as the number of person-hours available to do content moderation). Sahar Massachi, a former Facebook software engineer in Civic Integrity and one of the founders of the Integrity Institute, sometimes refers to this as the difference between adding more traffic cops and adding speed bumps to the roads.[8]

With respect to social media platforms, free speech ideals also place important constraints on the scope of potential government regulation. While few countries have free speech traditions quite as strong as those of the United States, the value

[8] Mostly in private conversation among those involved in the Institute, including myself; a less pithy presentation of the general idea is in Massachi (2021).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

is recognized in every liberal democracy and by international human rights law. Under US law, it would be impossible to directly restrain many of the pathological speech phenomena on social media (or to force the companies to do so), such as by banning the spread of conspiracy theories and so-called "fake news" or racist speech. Countries with hate speech regulations and the like, of course, might be able to go further, but even there – as with most efforts to regulate widely dispersed behavior via law – the practical difficulty of such regulations and the cost of enforcement is widely recognized. The same point holds with respect to transactional platforms. If, for example, states and municipalities really wanted to directly regulate every person who offered up a room on Airbnb, well, they could do so, but the effectiveness and practical scope of those regulations would be limited by the sheer number of people involved and the difficulty of the government directly monitoring so much behavior.

By contrast, regulations of the decisional infrastructure of technology companies themselves, as well as their design, and the affordances they offer – as the institutional approach recommends – are somewhat more insulated from free speech challenges insofar as they do not directly target expressive activity and could also apply to non-expressive activity. To be sure, regulations of companies are not without limits. In the United States, it would probably be unconstitutional to impose something like a "fairness doctrine" directly on social media companies, requiring them to engage in politically neutral content moderation (the misguided views of the State of Texas and the Fifth Circuit in a recent bill and case notwithstanding). It would also be unconstitutional, although somewhat more difficult to prove in a specific case, to impose regulatory disadvantage on them in ways motivated by political viewpoint, such as various Republican proposals to abolish Section 230 of the Communications Decency Act in retaliation for companies' interpretation of their rules to remove conservative speech from their platforms. But other sorts of institutional interventions are free from such constitutional barriers. For example, many interventions on the workplace structure of companies, such as a proposal given in the Conclusion to this book to give more job-related power to the offshored contractors currently serving as content moderators in social media companies, would not be directly related to speech at all. Also not directly related to free speech would be the use of antitrust law to break up companies with excessive market power (discussed later in this chapter). Similarly, government efforts to assist the voluntary building of user and public-dominated governance institutions for companies through industry-wide action like that proposed in Chapter 6 would be unlikely, because of their voluntary character, to run afoul of companies' First Amendment rights; for similar reasons, the interposition of an uncoerced choice by companies to build such institutions (albeit with the government's help) would at least arguably insulate the government's role from challenges on the basis of the First Amendment rights of users as well.

2.4 OBJECTION 1: IS PRIVATE GOVERNANCE AN OXYMORON OR A DANGER?

Before we delve into the account of how we might promote the effective governance of users by platforms, some interrelated preliminary objections must be addressed. The first identifies a kind of teleological mismatch between the goals of governance development and the ends of private, profit-motivated, entities. Even in the ideal form of the so-called "new governance" models under which the state recruits or collaborates with private entities to govern the public, there's an important sense in which the state remains in the driver's seat as much as possible – for example, by contracting with private entities to carry out governance tasks, recruiting the support of private entities for pieces of an overall project of infrastructure or community-building, or convening private for-profit entities to participate in their own regulation rather than the regulation of third parties.[9] By contrast, the most plausible ways to develop private platform governance involve platforms regulating their users, and, in view of the international nature of the major platforms, it's unlikely that governments will be able to exercise close supervision over the conduct of that governance. Indeed, close supervision by governments is almost certainly undesirable, due to the oppressive nature of some governments and the hegemonic nature of others – issues to be explored further below.

We have strong reasons to be wary of the delegation of governing responsibilities from states to private companies. At least in theory, states are oriented toward (and accountable to) the good of their people, whereas companies are oriented toward the good of their shareholders. In law, for example, the whole enterprise of "consumer protection" reflects a recognition that giving companies the power to dictate the behavior of people who engage in non-shareholder economic relationships with them (e.g., with excessively empowering contractual terms) makes individuals vulnerable to exploitation.

Even when states, rather than companies, have been the beneficiaries of capacity-building efforts, those efforts have often led to unintended consequences relating to the abuse of that power. For example, antiterror capacity-building programs in Kenya have also built the government's capacity to engage in repressive actions against local minorities (Bachmann and Hönke 2010). Similarly, with respect to companies, we have good reason to fear that increasing their control over the conduct of users on their platforms could facilitate any number of inappropriate choices

[9] Bell and Hindmoor (2012) give an extended argument for why these sorts of enterprises still depend on the notion of the state and its authority at the core. Arguably, Bell and Hindmoor's argument could apply to the program of platform governance development as well – we could simply understand it as a form of "metagovernance" in which states govern more effectively by arranging for some helpers. But in view of the fact that I explicitly propose institutions meant to dodge state control in a number of areas, particularly where authoritarian (or shakily democratic) states aim to turn platforms into cat's paws to do things like political censorship, I don't think I can avail myself of their argument.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

made by the companies over the years. For perhaps the most important example: Every platform company is widely criticized for violating the privacy of users; many scholars have suggested that such violations in the form of surveillance are the core business model of companies (e.g., Zuboff 2019). But by asking companies to do more governance we might be exacerbating their surveillance by giving it a kind of normative cover. After all, the same kinds of practices that offend our notions of privacy and raise the specter of surveillance might also be useful in effective governance. Consider database merging and user profiling across platforms – many of us object to, say, Facebook's using its infamous "pixel" to track our behavior across non-Meta-operated websites and even allegedly build profiles of users independent of their Meta-affiliated accounts (so-called "shadow profiles") (Garcia 2017; Aguiar et al. 2022; Bekos et al. 2022). But what if that information is used to identify "inauthentic" users, such as Russians attempting to subvert elections?[10] Suddenly, it seems perhaps a bit more acceptable, and perhaps even something we might demand of companies. Looking down the game tree, we may resist conferring additional governance resources (and responsibilities) on companies in order to avoid giving them further incentives and tools to conduct such surveillance.

This problem of surveillance – indeed, the general danger that capacity improvements for governance purposes (i.e., where company interests align with the public interest) may turn into capacity improvements for exploitative purposes (where company interests conflict with the public interest) – is real and serious.

However, I contend that we can avoid objections that point to the above-described problems by attending more carefully to the ways that (a) states and companies are subject to similar incentives (so state regulation of users directly, even if possible, might not be better than governance development); and (b) we can build capacity improvements for platform governance that do not necessarily empower companies as such – but rather empower other ecosystem actors with incentives aligned both with the public and with companies on governance issues – including users, nonuser citizens, civil society, workers and other important constituencies.

To begin, it is notable that many of our most useful accounts of state development suppose that leaders have self-interested motives, and, historically, it is difficult in the extreme to suppose that the nascent stages in the early development of the high-capacity nation-state were purely or even mostly public oriented. Indeed, a substantial part of the point of the literature on the development of governance

[10] At a Congressional hearing, Mark Zuckerberg denied any knowledge of "shadow profiles" but did claim that there were safety-related reasons to collect data on nonusers (Molitorisz, Meese, and Hagedorn 2021, 51), so this example is not hypothetical – although apparently the justification Zuckerberg gave was more focused on preventing third-party profile scraping (Hatmaker 2018), and hence on protecting privacy (I confess to some confusion about how collecting data on nonusers protects against privacy invasions by scrapers as such). More generally, Meta does use personnel and techniques from domains such as intelligence and criminal investigation to track down Russian subversion (Stamos 2018; Murphy 2019). So does Amazon (Amazon.com 2020).

capacity and the state itself is the design or evolution of institutions that can re-align the incentives of top-level and mid-level leaders to cause them to pursue the public good in their own private interest, or the identification of private interest based causal frameworks for the emergence of such institutions (e.g., Olson 1993; De Lara, Greif, and Jha 2008; North, Weingast, and Wallis 2009; sometimes this can include the development of private legal frameworks as well, e.g., L. Liu and Weingast 2018).

Unsurprisingly, the worries which I articulated above also appear frequently in the context of states. Take the problem of surveillance. It's useful to help states prevent crime, terrorism, subversion, and so forth. And it's incredibly dangerous to enable states to conduct political repression. I'd go so far as to say that the problem of surveillance in the context of states is the most classic and intractable example of the persistent tension between security and freedom. Shorter version: J. Edgar Hoover existed. Assuming that someone has to watch what people say on social media to control the spread of things like viral misinformation, there is little reason to think that the state is any less likely to abuse the power to do so than companies.

Moreover, it seems to me that the key problem with the objection to private governance is that it assumes a too-thick distinction between "public" and "private" which is, to some extent, merely a quirk of our place in history. Rather, we ought to understand that the "public," both in the sense of governments and in the sense of other ways of noncommercially organizing people (e.g., via civil society, social movements, and even purely artifactual organizations), can be integrated into the decision-making processes of private companies in ways that are *both* private and public – private because the companies remain for-profit businesses with private owners, whose decisions are in substantial part motivated by profit – yet public because those decisions can both be controlled in the public interest and improved in the private interest by integrating noneconomic forms of human organization, and hence human incentives, into them.

Critiques of private governance can also merge into critiques of the undemocratic nature of such forms of governance – and hence the implicit claim that what counts as democratic is limited by existing forms of public authority. As an example, consider two recent papers, one by Blayne Haggart and one by Haggart and Clara Iglesias Keller (Haggart 2020; Haggart and Keller 2021). The two papers criticize several major platform governance initiatives such as the Meta Oversight board and proposals by several governments, as well as leading scholarly proposals, on the basis of their lack of democratic legitimacy.[11] In particular, they focus on a conception of

[11] As so frequently happens in papers about "legitimacy," the authors seem to run together legitimacy in a normative sense, that is, decisions that ought to be respected because the right people made them, and legitimacy in a sociological sense, that is, decisions that will in fact be respected by relevant populations. For present purposes, I will assume that the core of their argument is about normative legitimacy, as that is what matters, morally speaking. (Sociological legitimacy might have moral importance as well, but only in the context of an argument according to which sociological legitimacy is required for normative legitimacy, or for some other important moral end. Otherwise, we're simply trying to get an "ought" from an "is.")

democratic deficit framed around what the Haggart/Keller paper describes as their lack of "input legitimacy," that is, the failure of major decisions to go through democratically accountable institutions.

There is force to their objections. For example, they object to the design of the Meta Oversight Board in virtue of the fact that its decisions ultimately are still in the control of the company, both because the company makes the rules and because the company sets the terms of civil society and other participation (e.g., by choosing the initial chairs). And, they criticize David Kaye's (2019b) human rights-focused approach for being overly concerned with protecting companies and their users against censorship efforts by governments, and failing to distinguish between democratic and authoritarian governments and the differing degrees of trust which ought to be extended to them in controlling online speech.

Yet the vision of those papers is limited by their rigid conception of the shapes democratic institutions might take. Both papers assume a fairly cramped Westphalian statist notion of democracy, in which it is impossible for a decision-making process to be legitimate unless it is associated with a specific, bounded, *demos*.[12] But if we value democracy because of the normative importance of individual and collective autonomy and the practical advantages of superior decision-making, I contend that those values can be achieved outside the context of the nation-state (back to the Deweyian point about how democratic institutions need to proceed from an understanding of the relevant underlying public and, well, its problems). Similarly, I think Kaye has the better of the argument with Haggart and Keller with respect to the danger of excessive control over platform speech even by nominally democratic states – we need only consider the threatened abuses of government power in weak democracies run by demagogues such as the United States during the Trump regime or India during the Modi regime to illustrate the risks involved. More generally, even robust democracies have long-term minorities, indigenous peoples, and other groups of people identified with, but not fully represented by, the nation-state that happens to rule the physical territory. For example, Spain is widely seen as a stable and legitimate democracy, but it is far from obvious that it can be understood to fairly speak for its Basque and Catalan citizens; how much less can the United States be understood to fairly speak for its Native American citizens.

It seems to me that the goal of achieving "input legitimacy" is a worthy one, but that we ought to focus our attention on ways to do so that include, but are not limited

[12] This is a consequence of the intellectual frameworks within which they operate: the Haggart paper relies on Dani Rodrik's account of the difficulty of democratic "national self-determination" in the global context, and the Haggart/Keller paper relies on Vivien Schmidt's claim that "input legitimacy" requires a kind of thick national polity with a discrete identity to which decision-making processes are accountable. Proceeding from those starting points, the arguments in those papers make perfect sense. But those starting points are controversial. (I happen to think they're outright false, but can't defend that claim in this book except insofar as some of the discussions of platform identity and constitutional patriotism in Chapter 5 might constitute a defense.)

to, the nation-state. Rather, this book will argue for platform governance reforms that can genuinely incorporate both democratic peoples organized into such states *and* national minorities, global civil society and social movements, and other forms of meaningful human organization not encompassed by the modern state. I will further argue that such reforms can actually make platform governance more effective at governance – that is, at writing rules that work and can be effectively enforced to control harmful externalities. The state, as well as platform companies, must drive such reforms, because it is states (and multi-state organizations like the EU) and companies that currently have the power to do so. But those empowered by the reforms cannot be limited to states and companies.

A better approach is the one described by Hannah Bloch-Wehba (2019, 67–71) in what I believe to be the closest affinity in the existing literature to this book's governance development approach. While Bloch-Wehba does not argue for capacity building as such, she does argue for the building of global democratic institutions to govern platforms as a form of what we might call legitimacy building, to contrast with capacity building. By doing so, she recognizes in a way that Haggart and Keller do not, that novel contexts for the exercise of power call for novel forms of democratization that are not wedded to pre-existing institutional forms.

The objection to private governance as raising a kind of teleological conflict between the interests of private companies and the public interest represented by states also ignores the scope limitation given at the beginning of this book. The scope of governance development as I propose is limited to the category of problems mentioned in the Introduction, where there is a *minimal alignment of interest* between platform companies and the rest of us. "Minimal" stands in for the recognition that platform companies might have complicated and conflicting interests, especially when considering the difference between short-term and long-term perspectives. "The rest of us" stands in, more or less, for the people of reasonably liberal and democratic states with interests in socially healthy behavior, an avoidance of fraud and malicious political subversion, and the like. The notion of minimal alignment of interests is also meant to recognize that sometimes the role of public policy is in part to give companies an incentive to follow the side of their interest which they share with the rest of us. Consider the problem of viral political misinformation: This is clearly against the public interest (poisoning democratic debate with lies is bad for everyone except the malicious actors spreading them). And it is *partly* against company interests – becoming known as a place where your quirky uncle Terry learns to be afraid of vaccines is not good for the long-term health of companies (or uncle Terry), as both users and advertisers are likely to be driven away. However, it is also partly consistent with their interests, understood from a different (more short-term) perspective: Viral misinformation is still activity, and it's activity that some users vigorously consume; in an economic model that monetizes engagement, toxic engagement still counts. Governance reforms might make a difference in those kinds of problems by giving companies the incentive and the ability to focus

on the long-run benefits of getting rid of the viral misinformation rather than the short-run benefits of allowing it.[13]

By contrast, the problems where such alignment of interest does not exist are not amenable to governance reform. Consider the surveillance problem again: critics of the platforms represent it, perhaps rightly, as a pure clash of interests between companies and the rest of us. They make more money by doing and facilitating more surveillance; we enjoy more privacy, freedom, and power over our own lives by being subject to less of it. These problems are serious and critical, and we need to find a way to address them – but they are not the subject of this book, except insofar as they are indirectly *implicated* by this book, for example, because we need to craft our governance reforms with an eye toward not making the conflict of interest between platform companies and the rest of us worse.

That being said, in an ideal world, well-crafted governance development might also change platform incentives in the right direction even with respect to those situations where there is a genuine clash of interests between companies and the rest of us. In particular, suppose I'm right that more effective governance – over issues where companies' interests are aligned with the rest of us – requires giving ordinary people more power within and over companies (as I argue in the rest of the book). Those same reforms might also give the rest of us enough influence to at least partly nudge company incentives over a little bit in places where they are *not* aligned. More bluntly, once we democratize the platforms, we might discover that our new democratic institutions can help them find revenue models that are less dependent on surveillance, the exploitation of workers, the evasion of local regulation, and all the rest of their abusive behaviors.

2.5 Objection 2: Does Platform Capacity Building Over-Empower the Already Powerful? Is It Colonial? Imperial?

Here in the United States, our problems tend to dominate not only our own but the world's minds. Because of our economic power, our ideologies also tend to dominate platform governance. Twitter's executives used to call it "the free speech wing of the free speech party," in line with America's position on the global extremes of free speech and the fact that, at the time, its leadership was dominated by famously libertarian Silicon-Valley types. On the other end of the libertarian spectrum, the same pattern holds: America's "leadership" (if we can call it that) in hyper-aggressive intellectual property regulation and the power of its media companies have exercised an outsized influence on global intellectual property law (e.g., Kaminski 2014; Yu 2006), doubtless including its enforcement by platform companies.

[13] At a sufficiently – and uselessly – high level of abstraction, we might say that the basic nature of any governance reform is to decrease the extent to which the ones doing the governing discount the future relative to the present.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

To a limited extent, Europe has also been able to propagate its values via platforms. Because of the size and wealth of its market, a number of aggressive EU internet regulations have exercised an international influence – the General Data Protection Regulation being the most prominent, but the Digital Markets Act and the Digital Services Act have also drawn particular attention, as have a number of proposed or implemented intellectual property regulations.

Accordingly, any interventions that might make these firms more effective at governing platform behavior potentially promote the further propagation of US and European values to people and places that may not want them. US free speech ideals might, for example, be applied in regulating conduct on social media in countries with much less libertarian norms; US and European intellectual property regulations might be used to censor the speech of users from countries and cultures with a very different attitude toward things like authorship and copying.

At the limit, Americans may even be making extremely contentious political judgments about the legitimacy of the leaders of other states. At least when Donald Trump was kicked off Twitter and Facebook, the executives making the decision were Americans, in American companies. But what happens if, say, Narendra Modi or Jair Bolsonaro were to incite a similar attack on their national legislatures after losing an election? Would we be as comfortable with the likes of Dorsey and Zuckerberg (or, heaven help us, Musk) making the same call?

Consider as a concrete example of the difficulty of these determinations the Thai laws against insulting the monarchy. There's a legitimate debate to be had about whether those laws – restrictions on political speech, to be sure, but restrictions that are associated with a longstanding political tradition and that are, at least on their face, fairly narrow – are consistent with human rights principles. Platform companies find themselves thrust into such debates when, for example, the Thai government asks YouTube to take down videos committing lèse-majesté (DeNardis 2014, 218; Klonick 2018, 1623).[14]

Exacerbating this risk of colonialism is the notorious problem of cultural competence experienced by many platforms, which frequently lack political, social, and even linguistic knowledge sufficient to adequately interpret behavior or model the values of foreign cultures. Perhaps the most famous example is Facebook's habit of taking down pictures of indigenous persons with culturally appropriate bared breasts (Alexander 2016), evidently on the basis of having misinterpreted the expression as pornography.

As with so many issues in internet governance, the problem of colonialism has not begun with the phenomenon of platforms. Rather, even governance strategies for the basic technical infrastructure of the Internet, such as technical standard-setting

[14] However, this example may be inapposite, insofar as the Thai lèse-majesté law is used for political repression, as Streckfuss's (1995) account suggests. On the history and social function of the law more generally, see Mérieau (2019, 2021). It is notable that Thai commentators have compared the law to the laws against insulting the Prophet in Muslim countries (Mérieau 2019, 57), a similar area of cross-cultural conflict about the function of free speech in the context of sacred or sacralized figures.

processes for things like IP addresses, WiFi, and the like, can easily exhibit features that could be described as having at least a semi-colonial character. DeNardis (2009, 201–5) describes how such standard-setting processes can become dominated by developed country politics, and, in doing so, embed economic advantages to developed countries, such as a preference for incumbent intellectual property holders in those countries rather than new entrepreneurs in developing countries.

Moreover, there is a distinctive historical fidelity to the worry about colonialism in private governance. A major driver of European colonialism as it actually occurred was the empowerment of companies like the Dutch East India Company, in a process that historical sociologist Julia Adams (1994, 329–34) described as "patrimonial" in virtue of the way that the Dutch state, under the influence of powerful merchant elites, granted the company legal and financial privileges and sovereign authority while structuring its operations in order to reinforce their own institutional power. Adams (1994, 335–36) argues that the organizational structure of the company facilitated colonial enterprises of military and economic domination up to and including genocide by providing a mechanism through which the state's repressive goals could be laundered.

The enterprise of rule-of-law development too has been criticized as colonial – including in my own prior work (Gowder 2016, 170–71; see also Mattei and Nader 2008). Rule-of-law development projects often appear to rest on the assumption that developing nations lack legal systems of their own – because they don't look exactly like the western ones – and attempt to impose the values and institutions of the countries – mostly the United States and Western Europe – doing the promotion on the nations in which such projects are carried out. Certain kinds of capacity-building assistance may be downright harmful to recipient countries, perhaps even imposed on them coercively. For example, the enterprise of capacity building has been deployed in order to "support" developing countries engaging in more vigorous intellectual property rights enforcement (May 2004) – even though local economies may benefit from a lower degree of intellectual property protection than the international norm, and even though the international norm might be understood as being imposed by the greater trading power of powerful countries like the United States. My suggestion that platform governance regulation amounts to a kind of capacity development similar to rule-of-law development in the state context might seem to invite similar worries.

The conduct of major platforms has already been criticized as colonial. For example, Kwet (2019, 12) criticizes Facebook's "Free Basics" internet service, which provides network connectivity for free in developing countries as a way to acquire market dominance via preferring the company's own services, in view of the way such market positions enable such companies to exercise control over communications in those countries and thereby, in Kwet's (2019, 12) words, "undermines various forms of local governance." Similarly, Cohen (2019, 59–61) interprets platform extraction of data from and surveillance of the global South – sometimes with the collaboration of local governments – through a colonial lens.

Moreover, other scholars have identified the dominance of US platforms with the concept of imperialism. As I read Dal Yong Jin's (2013, 2017) influential work on the subject, platform imperialism is primarily economic in nature, and revolves around the way that platforms leverage intellectual property law as well as the commodification of information created by as well as about their users to extract wealth from the global South. In this sense, it can be distinguished from colonialism, which I mean to refer to political and social rather than economic domination – that is, the undermining of the freedom of colonized peoples to determine their own fate. However, the two phenomena obviously tend to go together – the propagation of American (and European, etc.) cultural influence increases both American economic influence and American social/political influence, and the two sorts of influence can be self-reinforcing.

This imperial quality can also be a consequence simply of the use of platforms as a kind of standard due to their positive network externalities. As David Grewal (2008) has cogently argued, positive network externalities constitute a form of power in that they degrade the usefulness and availability of alternatives, and hence deprive people of access to those alternatives. Even if nobody is in control of the object of one of these kinds of positive network externality coordination goods (or "standards"), such as a language or a currency being used as a lingua franca, Grewal's work shows that nonetheless, they have the potential to suppress diversity and local autonomy. So, for example, the network externality-derived usefulness of Facebook or Google in the context of a global economy may impede the development of social media or search engines that are more responsive to local needs.

Another form of colonialism or neocolonialism (quasi-colonialism?) in the platform business revolves around their labor practices. In particular, Roberts (2019, 182–200) describes how the commercial content moderation enterprise is embedded in colonial relationships, as offshored content moderators in the Philippines work in their (exploitative) jobs under conditions both of colonial cultural influence (this work is done in the Philippines partly because American colonial legacies have left workers there with a high degree of cultural and linguistic competence relative to American culture) as well as company-driven development paths in things like infrastructure choices.[15]

Objections to private governance and objections to colonialism potentially come together, as one tool with which developed nations have objectionably dominated developing nations has been the integration of nonstate organizations into government. Even leaving aside historical examples such as the various East India companies, contemporary governance capacity-building efforts have been criticized for effectively replacing functions of local states with private organizations that appear

[15] One of the key recommendations of this book is more substantially integrating such workers in company decision-making, a recommendation which may directly alleviate this particular colonial aspect of those relationships.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

to answer more readily to the donor states than to the domestic governments they are supposed to serve (e.g., Mkandawire 2002, 16–18).

In a world in which every government was legitimate and just, some potential solutions would suggest themselves, primarily involving platform rules being modified and applied in accordance with the requests of local governments. But, of course, this is not the case, and, in fact, the quasi-colonial character of platform governance is actually valuable to many living under authoritarian regimes, particularly with respect to social media. For example, one of the stories we might tell about the role of social media in the Arab Spring is that revolutionaries and dissidents took advantage of the fact that companies like Twitter imposed American free speech norms on content in countries like Egypt. For a more recent example, Brazilian populist demagogue Jair Bolsonaro attempted (albeit unsuccessfully) to prevent social media companies from removing his COVID-19 misinformation (*BBC News* 2021; Marcello 2021). Thus there is an inherent tension between the goal to promote the self-determination of peoples with diverse cultural and value systems and the goal to not be a tool of government oppression or demagoguery.

Nor can international human rights standards, on their own, resolve the tension between platform colonialism and platform culpability in governments gone bad, for human rights standards, like all laws and other normative principles, are subject to interpretive disputes, and it may still be objectionable for Western elites to make themselves the authorities to decide whether, for example, the request from some government to censor some content from one of its citizens comports with those standards. Moreover, the content of those standards themselves is disputed, especially by states that see themselves as having a religious identity shared with the entire community – thus, for example, the Cairo Declaration on Human Rights in Islam is widely seen as a response to the Universal Declaration on Human Rights which articulates a contrasting framework meant to be specifically Islamic.[16]

Even among wealthy liberal-democratic nations, there is disagreement on the scope and applicability of human rights norms, and that disagreement has already leaked into the platform governance domain. Internet human rights group Article 19 has criticized Germany's "NetzDG" intermediary liability law for (among other things) imposing liability on companies who do not help enforce Germany's prohibition against glorifying the National Socialist party – a law that, in Article 19's view, violates international human rights norms with respect to free speech.[17] To me, Article 19's objection is deeply misguided: Germany has prohibited the National Socialist party for a very long time and *we all know why*. It's hard to imagine a more

[16] A later human rights instrument with many of the same signatories, the Arab Charter on Human Rights, has been criticized for itself violating international human rights standards, in particular, in its denial of the principle of self-determination with respect to its condemnation of Zionism (Barnidge Jr 2018; United Nations 2008).

[17] Article 19, "Germany: The Act to Improve Enforcement of the Law in Social Networks," August 2017, www.article19.org/wp-content/uploads/2017/12/170901-Legal-Analysis-German-NetzDG-Act.pdf p. 15.

compelling case for deference to a local interpretation of a global human rights norm in light of a country's distinctive history, culture, values, and dangers. But this just goes to show that even the extent to which global human rights standards are subject to local variation is itself controversial.

2.5.1 *Mitigating Platform Neocolonialism*

The force of the neocolonialism objection may be limited in two respects. First, it does not apply to platform governance innovations that are focused on the sorts of wealthy Western countries which are traditionally the sources rather than the victims of colonialism. Consider the governance failures that social media companies' inability to deal with the American alt-right and the incitement that led up to the January 6 attacks revealed: US-focused governance remedies that apply only in the United States, to be implemented by other Americans, might be lots of things – potentially racist and sexist, for example – but they are unlikely to be objectionable on grounds of colonialism (with the potential exception of disagreements about how such innovations are to be applied to situations involving Native American nations).

Second, to the extent objections to capacity-building efforts rooted in *uncontroversial* applications of international human rights principles (e.g., the prohibition of genocide or states repressing nonviolent opposition parties) are rooted in objections to liberal democracy itself, it may be that such objections are less to capacity building and more to the very presence of platform companies at all. As I argue in more detail in Chapter 5, there's a sense in which the platforms are inherently oriented to liberal democracy.[18] If countries don't wish to comport with minimum human rights standards, they can always prohibit their citizens from using the platforms in question, as China does. In view of the fact that platforms are not imposed on countries, if a country rejects the very idea of human rights itself it seems to me that it is fully satisfactory to say that doing so amounts to a rejection of platforms as inherently encapsulating ideas like free markets and a discursive public sphere.

However, for the reasons discussed above, the foregoing is not a full answer to objections to human rights-derived platform governance. Countries or peoples originating in different cultural contexts may accept the idea of human rights but have different legitimate interpretations of the contents of human rights, or of the relationship between the behavior of individuals, companies, and states to human rights standards or other relevant normative standards. The problem of free expression is particularly salient here, not merely because of the German example noted above but also because the major platform companies originate in the United States, a country that is widely recognized as a global outlier in its free speech absolutism. And while recent developments (such as Musk's welcoming the alt-right back to

[18] For more on rooting content moderation decisions in company values, see Bruckman (2022, 205–7).

Twitter) are extreme even by American standards, the people of other countries have previously noted a kind of American bias in platform interpretations of free speech norms relative to the interpretations of other countries relating to matters such as hate speech. In the words of one "non-Anglophone European" content moderator for Facebook, as told to the New Yorker "If you're Mark Zuckerberg then I'm sure applying one minimal set of standards everywhere in the world seems like a form of universalism[.] To me, it seems like a kind of libertarian imperialism, especially if there's no way for the standards to be strengthened, no matter how many people complain" (Marantz 2020).

Part of the answer to the problem of platform colonialism from capacity-building efforts may just be to observe that the status quo ante is already terrible. Americans and (to a lesser extent) Europeans are *already* regulating the behavior of people in other countries and imposing their own values on them. It's not as if the alternative to capacity development for platform governance is autonomous self-governance by people from all countries and cultures, or even a void of governance. Likewise, countries such as the United States are already using platforms as an instrument of their own foreign policy and have the potential to do so to a greater extent under existing law and company institutional structures. For example, United States Customs and Amazon have run at least one "joint operation" to control the import of counterfeit goods (U.S. Immigration and Customs Enforcement 2020), and Amazon has reportedly channeled counterfeiters to police agencies in numerous countries (Brodkin 2021) – in effect helping the United States enforce an international as well as domestic intellectual property regime that, of course, is fairly heavily weighted toward US interests and against the interests of developing countries who may benefit from looser enforcement. On the other end of the platform economy, there has been substantial speculation that social media companies could be held liable under American "material support" statutes for hosting content by organizations that it designates as terrorists (VanLandingham 2017; Bedell and Wittes 2016).

Rather than a world free of platform colonialism, the alternative to governance reform may simply be destructive incompetent colonial governance partly by ham-fisted corporate efforts to maintain the market appeal of their platforms, partly by the unintended consequences of product design for other purposes, and partly by the short-term political goals of leaders in powerful countries – much like we currently have. Genuine governance reforms directed at platform companies may make things better almost by default.

Some kinds of governance improvements are also inherently improvements from an anti-colonial standpoint. For example, one of the key recommendations of this book is the integration of nonstate groups in both more and less economically developed countries into platform decisions. Such processes would track emerging human rights standards; for example, Article 19 of the United Nations Declaration on the Rights of Indigenous Peoples provides that such peoples should be consulted on decisions that affect them. To the extent we understand that the right to apply to

platform companies, including indigenous peoples in company decision processes represents a direct implementation of that right; in countries with governments directly descending from settler colonialism (such as the United States, Brazil, Australia, and Canada, for the most obvious examples), including indigenous peoples in company decisions, where previously only governments largely representing non-indigenous peoples had any influence over them, would amount to a positive reduction in colonialism.

That being said, these points do not eliminate the reason to guard against colonialism in platform governance development proposals. Even if any governance development program will represent an improvement over the (terrible) status quo, there can still be more or less colonial programs, and we in countries that bear culpability for the evil of colonialism and have the power to implement platform governance development programs have compelling moral reasons to implement those programs with the aim of reducing colonialism. Accordingly, the danger of colonialism in both cultural and political respects ought to be considered in any effort to reform and advance the capacity of platforms to govern their users. And we can state a minimal criterion for a successful reform proposal that it decrease, relative both to the status quo baseline and to other viable options, the risk of the United States and other wealthy powers, as well as corporate personnel aligned with those powers, exercising colonial power over the people of the global South.

In addition, governance improvements may exacerbate imperialism in the economic sense described above even if they do not exacerbate, or even if they mitigate, colonialism, insofar as they increase the capacity of platform companies to extract wealth from less economically dominant countries and peoples. Accordingly, governmental action to incorporate non-US peoples into platform governance should also be paired with efforts to bring about fair compensation for informational resources extracted from those peoples. However, while important, this is not the core concern of this book, as existing philosophical theories of global justice already offer a critique as well as proposed remedies for global economic imperialism (e.g., Caney 2001).

2.5.2 *Facebook Probably Shouldn't Have Been in Myanmar at All*

There's one easy way to avoid colonialism: It may be that American (and secondarily European) companies simply ought not to be in some countries. If companies lack the cultural competence or legitimacy (in either a sociological or a normative sense) to make rules for user behavior without creating injustice or exacerbating existing harms in a given country, and if the government of that country cannot be trusted to give the company its marching orders, then we might simply conclude that the presence of a given platform in a given country may cause more harm than good, and it ought to withdraw from the market.

In principle, the governments of developed countries – particularly the United States – in conjunction perhaps with United Nations human rights agencies, could

implement legislation in accordance with the previous paragraph. The United States already has a substantial legal framework permitting the designation of countries with which its companies are not permitted to trade. And platform companies are already covered by some of those laws, or at least appear to in some cases be voluntarily complying with them in the absence of debates over coverage – for example, there is evidence that Facebook has made use of the US government's "Specially Designated Global Terrorists" list in developing its dangerous individuals ban list (Biddle 2021), which may amount to compliance with individually targeted sanctions provisions against those individuals.[19] In principle, legislation could be enacted prohibiting any US company from operating in a country with respect to which the State Department or the United Nations High Commissioner for Human Rights has certified that the company in question has a severe, ongoing, negative human rights impact. And while the United States First Amendment may pose some challenges to the application of such a rule to social media, it may be defensible, particularly if the legislation in question is not specifically targeted at speech and applies to many other US companies as well. Such breadth, however, likely makes any legislation along these lines politically impossible: consider its application to US resource extraction companies which have generated numerous terrible human rights impacts, and the lobbying power of such companies (e.g., of possible application, see Giuliani and Macchi 2014, 485).

2.5.3 *Dispersing Power: Simultaneously a Governance Reform and an Anti-colonial Measure?*

Beyond outright withdrawal, another strategy for avoiding colonialism in a context in which the American state as well as American (and secondarily European) cultural values currently exercise a substantial governing role over people in less developed countries, is to design governance reforms that *reduce* and *disperse* that power, not to other Americans (or to local elites allied with particular elements in the American political/economic system), but to the actual people in the countries in question. In the later parts of this volume, I argue that dispersing power to the actual people whose interests are at stake (as opposed to a bunch of Americans) is actually likely to improve the efficacy of platform governance as well as to make it less colonial.

In the international rule-of-law development field (the area with which I am most familiar), similar goals have been integrated into some projects which have thereby avoided, or at least reduced, the danger of neocolonialism. Such projects have been distinctive for focusing on local empowerment: identifying sociologically legitimate

[19] See, for example, 31 CFR §§ 594.201, 204, 310, describing prohibition on providing services to specially designated global terrorists, although exceptions for "informational materials" and related items may arguably cover some Facebook services. I hedge with the "may" in the text because I lack expertise in sanctions law.

local leaders, including those independent of formal state institutions, and giving them the tools they need to participate in local dispute resolution.[20] Such projects have succeeded, not by attempting to impose American or European conceptions of law and of adjudication but by consciously accepting the legitimacy of legal pluralism and so-called "informal" adjudication. By, in effect, meeting people where they are, such projects simultaneously are (at least potentially) more successful at actually bringing the benefits associated with a legal order to the recipients of development assistance while reducing colonialism.

A first pass solution to the problem of platform colonialism may also involve similar forms of local empowerment. Rather than assuming that governments – which might be dictatorial, or simply unrepresentative – are the right entities to introduce local variation in either the content, interpretation, or enforcement of platform policies, it may be possible to turn to, for example, civil society leaders, religious leaders, activists, and other representatives of diverse local groups to carry out these functions. And while identifying the appropriate people to whom to turn and recruiting them into the governance framework can be challenging, again strategies from international development, such as hiring experts with deep local knowledge and social science training to identify relevant leaders, might be imported into the platform context.

However, there's still a degree of colonial imposition in such a plan. One of the things that colonizer powers actually did in the existing history of colonialism was identify and elevate local leaders whom the colonizers perceived, because of their own biases, preferences, and interests, to be most suitable. Some of the grimmest consequences of colonialism have their roots in this history; most infamously, it has been argued that the Rwandan genocide was at least in part attributable to the decision of Belgian and German colonizers to promote the Hutu/Tutsi ethnic distinction and make use of the Tutsi as a local aristocratic class (Newbury 1998). Obviously, similar practices must be avoided in the platform context. Fortunately, one advantage of the current time as opposed to the period of active colonialism is that there is an active transnational human rights community that can assist in identifying civil society organizations capable of participating in governance institutions, where human rights organizations – while doubtless still aligned with the interests of the powerful – are at least *more* neutral than companies and governments themselves.

Ultimately – and this is a point to which I will return in Chapter 6 – the only real answer to this problem is to accept imperfection and adopt a system of iterative inclusion, in which those who participate in governance institutions are subject to continuing critique and modification. The real challenge in institutional design is to create sufficiently open-ended modes of inclusion, which can receive feedback from excluded groups and respond to that feedback by effectively incorporating new people into decision-making processes, as well as structures for dispute resolution

[20] See discussion in Gowder (2016), 171–175.

between different groups who have a stake in any given decision.[21] This is all unhelpfully abstract right now, but this chapter is directed simply at describing the general character of the problem – concrete elucidation of the sorts of designs that might satisfy these criteria will have to await Part III of this book.

Moreover, the attention to mitigating colonialism must also include attention toward mitigating other injustices, and sometimes there is a trade-off between the two – or at least an apparent trade-off. Consider the problem of subordinated persons within colonized peoples. For example, Bina Agarwal (2001, 1629) observes that when women are included in local forestry groups in India, it is often because of external pressure by "a local NGO, forest official, or international donor." When women are included, resources are managed more effectively – but is it colonial for international development "donors" to impose the inclusion of women on local institutions which traditionally operate under more patriarchal norms?[22] Perhaps, but there is a moral trade-off to be made (as well as an effectiveness tradeoff, since Agarwal also observes that gender-inclusive forestry groups work better), and ultimately I shall argue that the goal should be to include everyone – states in the global South *and* the indigenous and colonized peoples whom such states often inadequately represent *and* those, including often women, whom indigenous and colonized peoples themselves often inadequately represent. The only way to do so, as I said a moment ago, will be via an iterative learning process in which claims for inclusion by those who are left out of any pre-existing institutional arrangement are sought out and prioritized for action.

However, it may be that solutions to injustices within colonized peoples can only emerge from the peoples themselves. This is how I read Fanon's (1994, 35–67) famous essay on the veil: while veiling may be an injustice to the extent it is imposed by tradition on (as opposed to voluntarily chosen by) women in colonized cultures,[23] a reconciliation between the autonomy of a colonized people and the remedy of the injustices within a pre-colonized culture can only occur through (and after) the process of shaking off the chains of colonization, and from the self-assertion of the colonized.[24]

Particularly important in Fanon's account is the way the veil becomes operationalized by the oppressed in the course of the collective action of the Algerian revolution.

[21] If we believe Dewey (1927, 198–99), speaking of education but with logic that applies to many other social institutions, this is in a sense characteristic of all human social design: Using our knowledge of human sociality to intervene on the social world changes the underlying facts, and hence changes the interventions we may wish to make. Similarly, developing inclusive institutions, if done right, changes who is making the decisions about the shape of those very institutions, and hence changes who ought to be included. It's iterative all the way down.

[22] Cf. Okin (1999).

[23] I confess to not being sure about Fanon's view on the pre-colonial injustice of the veil.

[24] Moreover, it is easy – because of the psychology of the colonizer – to represent efforts to destroy the colonized culture as efforts to "free" those whom it allegedly oppresses. If that's right – and it seems right to me – then a project such as this book can never be untainted by colonialism. Even the judgment of which precolonial practices are unjust is likely to be inaccessible except from within colonized cultures.

Originally unveiling is a tool of the colonizer, but when women begin to participate in armed resistance, it becomes a tactical tool of that resistance, to be dropped or adopted as necessary in the broader goal of liberation. What this illustrates to me is that the important fact is the autonomy of those alleged to be oppressed even within a colonized people – autonomy that of course is necessarily conditioned both by the system of colonialism as well as by the restraints imposed by the colonized culture, but that can nonetheless be *taken* by their own self-assertion. Thus, returning to Agarwal's account of women's participation in the forest management process, she also notes that women had been observed to seize their own forms of participation through parallel organizations which were neither imposed on their communities from the outside nor merely licensed by the men (Bina Agarwal 2001, 1629–30). While she notes that these organizations are suboptimal as a method of including women insofar as the "official" organizations have formal control of the underlying forests, I submit that it might be a foundation for a kind of postcolonial self-assertion along Fanon's lines.

For present purposes, this suggests a modification to the concept of iterative inclusion, in which the ultimate goal of governance reforms is not to be understood merely as participation by colonized peoples but as *self-determination*, at least to the extent that such a thing is possible within the limited field of action encompassed by the book (for true self-determination would require changes well beyond the scope of the platform economy). That self-determination should be capable of interacting with any novel governance structure in both causal directions – that is, governance structures should be open to receiving demands of self-determination generated from the outside, but they should also be organized so as to promote the implementation of demands for self-determination generated from the inside.[25]

2.6 Objection 3: The Whole Industry Is Bad – Radically Remake It by Force

One question that immediately arises as we consider the harms platform companies have caused is: Should we (as citizens of democratic states) simply put a stop to them? Should we ban their services (consistent with constitutional and international free speech and other limitations)? Should we tax them into oblivion?

Those who write about the major platforms almost universally emphasize the social harms they create. And, doubtless, those harms are significant. Amazon crushes small businesses and abuses its workers, while facilitating a seemingly endless supply of counterfeits and scams. Facebook and Twitter and Instagram and the like promote viral conspiracy theories, political polarization (maybe, the empirical

[25] This issue may be intrinsic to human rights in general. Abdullahi Ahmed An-Naim (2021) makes an argument along these lines, suggesting that state-centric conceptions of human rights entail either relying on the very states that are violating those rights, or relying on external coercive intervention which itself creates human rights violations – instead, the protection of human rights should arise from political and social movements.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

jury is still out (Tucker et al. 2018)), and psychologically unhealthy forms of social self-presentation and "news feed" addiction. Airbnb and Uber and the like make it dramatically more difficult for us to democratically regulate industries like hotel-keeping and livery services, and have engaged in troublingly exploitative work practices under the guise of independent contracting.

But Facebook and (pre-Elon) Twitter and Instagram have also provided many of us the services listed on the tin: They have allowed us to maintain and build real connections with other people which we might not otherwise have maintained.[26] There are friends from the many places I've lived with whom I keep in touch almost exclusively over social media; and the fact that the pictures of their dog or kid or whatever occasionally scroll through my feed makes it much easier to maintain the rewarding social relationships. Those relationships have social and emotional value that is real, and that was much harder to maintain before the existence of social media.[27] Those relationships have economic value too: as has been well understood at least since Granovetter (1973), things like jobs and other valuable opportunities are often learned about through "weak ties" – that is, people with whom one is acquainted, but less close than, say, one's intimate friends or the colleagues whom one sees daily – precisely the ties that social media most helps us maintain.[28]

[26] One caution that must be given is that revealed preference theories of social value cannot do the work here. A naive defense of platforms would take the fact that people choose to use them in the face of alternatives as evidence that they provide a benefit. For example, one study suggested that people would require compensation of just under fifty dollars a month to give up Facebook (Brynjolfsson, Collis, and Eggers 2019). The problem is that revealed preference theories of value are not sound, because people's choices are endogenous to the market in which they find themselves. Thus, many critics of social media rightly worry that these products are addictive, and that people would be better off if they were protected from the urge to use them – a supposition supported by a number of empirical studies (e.g., Tromholt 2016; Dhir et al. 2018; Hunt et al. 2018). We also know that the entrance of new options in a market can affect the available alternatives – a prominent example being the way that social media engagement algorithms have degraded the quality of journalism by making available to people articles which reward attention to headlines rather than those which reward critical attention to content. (For a helpful discussion of endogenous preferences in markets, see Satz 2010, 180–81.) Finally, of course, we know that platforms, like all other forms of social activity, are sometimes characterized by negative externalities: individuals may personally benefit from their activity on platforms, but those activities may harm uninvolved third parties; revealed preference theories of social value only capture the benefit to participants and not the harm to bystanders.

[27] Of course, not every social media platform provides a meaningful amount of social value. It's unlikely that Parler and Gab create any benefit to society, due to their function as focal points for right-wing extremist gathering. In a more nonpartisan context, many hyper-specialized kinds of social media can be hyper-specialized in things that are socially pernicious. The most striking example as of this writing is a downright dystopian application called "Citizen," which is essentially a many-to-many social crime reporting platform that drives engagement by whipping up people into irrational levels of fear with pervasive "incident" notifications (that can be anything from an assault or a fire to a mere crowd of people), and, unsurprisingly, facilitates racial profiling as well as risking (or sometimes downright attempting to promote) vigilantism (Cox and Koebler 2021).

[28] There are sound reasons for this: people with strong ties are likely to occupy many of the same social spaces as oneself, and hence to learn about the same opportunities that one already is aware of. Weak ties are likely to bring to one's attention opportunities arising in social contexts that one does not occupy.

And, indeed, I have learned of important publication and consulting opportunities from weak ties via social media.[29]

The same can be said for the transactional platforms. For all the harm that businesses like Uber and Airbnb have caused, the fact of the matter is that they meet (or at least met at the time of their creation) a real need. For perhaps the most obvious example, the taxi market was egregiously anticompetitive if not outright corrupt until Uber and Lyft came along. At one point, New York taxi medallions were selling for up to 1.3 million dollars (Harnett 2018). That's blatantly corrupt: drivers would go into life-ruining levels of debt (Rosenthal 2019) to buy the privilege of participating in an artificially supply-constrained market; not all that different from student loans to go to medical school or law school, except that those professional degree programs provide an actual education necessary to do the job, whereas the taxi medallion system was just a pure government-sponsored cartel.[30] Or they would have to rent their medallions from those who could afford to own them – so that the government monopolies gave way to investor monopolies in which a rentier class who (until Uber and Lyft came along) effectively took no risk (thanks to the artificially constrained supply) and could simply print money.[31] What possibly legitimate policy reason could there have been to create a system in which investors could speculate on occupational licenses? The level of total policy failure represented by the medallion system is just astonishing: by creating an artificially supply-constrained cartel and then putting membership in that cartel on the market, it managed to combine the worst aspects of capitalism with the worst aspects of central planning. Uber and Lyft did a good thing by blowing it up; the taxi drivers who were harmed in the process should seek recompense from the corrupt politicians and medallion oligarchs who built the predecessor system in the first place.

Similarly, Amazon, for all its brutal market power, in all likelihood helped save lives during the COVID-19 pandemic as its massive and painstakingly built logistics network meant that it could actually deliver necessities to people at semi-affordable prices rather than send them out into retail stores to infect one another with a deadly disease (and yet, at the same time, they probably killed a few as well by providing a platform for fraudulent sellers to distribute counterfeit N95 masks). Similar claims could easily be made of food delivery companies like Doordash and Instacart. In the beginning of 2021, as the federal government struggled to manage the distribution of the COVID-19 vaccine, President Biden reportedly considered recruiting companies like Amazon to help – and rightly so, since the kind of pervasive logistics

[29] My experience consulting with Facebook itself came from a then-weak tie who worked there, and who announced that her team was looking for someone matching my profile over Facebook (although ultimately working together for a period transitioned us from weak to strong ties).

[30] Even if you accept the so-called "signaling" theory of education, at least education actually provides that signal – whereas taxi medallions were nothing at all but an artificial supply constraint.

[31] For example, one investor in 2006 bought an entire block of city-released medallions at auction (Olshan 2006).

network that can somehow reliably deliver a potholder to the door of almost any American in two days had the potential to be immensely helpful in getting vaccines into arms (Scola and Ravindranth 2021).[32]

The big mainstream platforms also cause undeniable harms. I take no position on whether the overall balance of social benefits and harms caused by platforms is in the black or the red. But I wish to urge caution here: nobody else is likely to be able to carry out that balancing act either. The harms they have inflicted on the world are probably incommensurable with their benefits, and that evaluation embarrasses the case for taking a prohibitionist approach to the companies.

For example: During the COVID-19 pandemic Amazon, Instacart, and similar delivery companies helped protect the health of those who could afford their services at the expense of workers, such as warehouse employees and delivery "contractors" (really, misclassified workers), who were unjustly placed at greater risk for the virus without being fairly compensated for taking that risk or, in many cases, having any real choice about the matter. This is a moral stain on the companies, on everyone who benefited from the harms inflicted on those workers, and on society at large. It is also a fairly common experience in capitalism, no different in any meaningful sense from the moral wrong committed for decades when Americans benefited from cheap and readily available energy to drive economic growth on the backs of brutally endangered coal miners. Critics from the left have been pointing out that workers are exploited under capitalism and endangered for the benefit of others for centuries, this has nothing to do with the platform economy. But can we honestly say that the world would be better off if Instacart had not existed during the pandemic? Or is the genuinely honest thing to say a variant of "something like Instacart should exist, but combined with massive redistributive taxation to ensure that those who benefit fairly compensate those who take the risk, and those who take

[32] At this point, I suspect that some readers will accuse me of peddling "technology solutionism." But I fear that the concept is easy to misuse. Sometimes technological solutions – or at least improvements, mitigations to problems – actually do exist. For example, Wymant et al. (2021) estimate that several hundred thousand COVID-19 cases in the UK over a three-month period were prevented by the National Health Service's contact-tracing app, built on a joint Apple/Google API. If those estimates are correct, then – subject to potential (but hard-to-estimate) caveats about alternative ways to have expended the resources that the UK government spent on app development and perhaps sociological effects according to which contact tracing apps undermined the impetus for more traditional and effective public health measures – real human lives were saved, and that is unquestionably morally valuable. At its best, the concept of technology solutionism through the insights of thoughtful (if sometimes polemical) commentators like Evgeny Morozov (2013) serves to draw our attention to the risks of oversimplifying complex social problems and imagining that simple technological changes can fix them without attention to their underlying causes or networks of effects. When deployed that way, the critique of technology solutionism is compatible with the argument raised in this book (which, after all, directly calls for the incorporation of messy humanity and complexity into the governance of internet platforms). But sometimes "solutionism" can be used to stand in for the denial that technological innovations can improve anything, or for the claim that technological changes contribute no social value at all – and in that form, I think it's just a mistake.

the risk do so freely?"[33] And in the absence of an obvious route to that alternative, does it make a lot of sense to try to engage in global evaluations of the with-Instacart and without-Instacart worlds?

Again, the same is true of the social platforms. Facebook and Twitter and Reddit and the like have unquestionably facilitated immense harms to our societies. Many people plausibly believe that Donald Trump wouldn't have been president without them, and that a mound of corpses of unknown size (related, e.g., to mismanagement of the COVID-19 pandemic and brutal immigration policies) is thereby attributable to their existence. Maybe so, but none of us have any clue how to actually evaluate that counterfactual in a world in which, for example, Fox News and talk radio exist, and in which the existence of the Internet itself makes it almost inevitable that spaces for dangerous political groups will exist outside the scope of big technology companies.

At any rate, heavy-handed regulation is a notoriously ineffective strategy for controlling products that produce real benefits to their users, even if they also produce social harms. Consider Uber again. In principle, municipalities could ban un-medallioned transport-for-hire altogether, and could subject people who offered or used those services to criminal sanctions. But this kind of prohibition strategy as to products and services that are already popular does not, it is fair to say, have a terribly good record. In all likelihood, now that the ridesharing genie is out of the bottle, any such prohibitory efforts would both increase the pervasiveness and harms caused by the intrusion of law enforcement in day-to-day life (you think the corporate surveillance done by platforms is bad? imagine the government surveillance necessary to control ride-sharing!) and be unsuccessful at controlling the conduct; in fact, exacerbate the harms of the conduct by driving it to black markets. See, for example, alcohol prohibition, the war on drugs, anti-prostitution policing, and countless other examples.

2.7 A CAUTIONARY APPROACH TO ANTITRUST LAW

A major policy lever that both the American left and the American right seem eager to pull – ranging from Elizabeth Warren (Kolhatkar 2019) to Josh Hawley (Bartz 2021) – is the use of antitrust law to break up the platform companies on the order of Standard Oil or AT&T, or at least to prevent additional corporate acquisitions. In some sense, this is the opposite of a capacity development strategy, as the goal becomes not to improve companies' capacity to govern user behavior on platforms but at least partly to defang it – and defanging it is at least part of the reason for such proposals (certainly it motivates Hawley's, as the US political right have often decried companies' alleged censorship of their views – more on this in Chapter 4). Nonetheless, to the extent that some of the social harms from platforms come from their market dominance, it is worth a brief discussion of this alternative strategy to close this chapter.

[33] On the unfree character of terrible jobs, see Gowder (2014a).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

The premise of the use of antitrust law against platforms is that many of their pathologies are the consequence of market power: There are no realistic alternatives to Google for search or Facebook for social media (although surely the latter is plainly untrue), and hence the bad decisions of those companies are amplified across society at large, while network externalities make it unreasonably difficult for market discipline to control those decisions.

The political relevance of what many platforms – especially social media platforms – do adds fuel to the case for their breakup. As Tim Wu (2018, 132) has suggested, there may be a democratic case for increasing competition in the companies that control substantial shares of the public sphere similar to the case for efforts against traditional media concentration.[34] This is certainly a point that would be welcomed by the American political right, as there is increasing reason to think that at least Facebook intervenes (however ineffectually) to stop far-right movements from getting a foothold on the platform (Horwitz and Scheck 2021) – and, alas, in the United States, there appears to be very little distance between the extreme right and the mainstream Republican Party. Similarly, Zephyr Teachout (2020, 43–44) deploys the fundamental premise of this book – that platforms are behaving like governments – to suggest that we ought to see companies like Meta as in effect competing against the democratic authority of the United States, and hence as particularly amenable to being broken up.

However, conventional strategies of corporate dissolution may, unless paired with other substantial policy interventions, entail the sacrifice of important public interests that are actually served by the companies in their present forms. With respect to Amazon, notwithstanding all the company's serious misbehavior, it is eminently reasonable to suppose that there are genuine economies of scale in a national logistics network of the kind that Amazon built.

With respect to social media, some breakup strategies seem to risk – perhaps deliberately, as in the Hawley case noted above – undermining the goal of governance to the extent they might encourage companies that serve niche markets, which, in turn, may cater to socially harmful interests. Consider that, for all the harms they inflict, Meta and Alphabet have strong incentives rooted in their scale to be compatible with mass consumption markets: If, for example, either Facebook's News Feed or YouTube's recommended videos became pervasively filled with Nazi-adjacent content, both large numbers of ordinary users and advertisers would likely flee the platforms. The slow-rolling destruction of Twitter after the Musk acquisition, with users and advertisers fleeing the platform as the extreme right is welcomed back, stands as a clear example.[35]

[34] This is in addition to the general democratic case articulated by Wu and others for breakups of companies that exercise excessive political power through the ordinary means of wealth and influence.

[35] While it is in principle possible for companies like Meta and Alphabet, aided by their industry-leading machine learning expertise, to segment their user and advertiser markets sufficiently fine that they can match users who are disposed to Nazi-adjacent content with that content and with advertisers who

By contrast, consider the actual niche social networks that have arisen in recent years: companies like Gab and Parler outright designed to appeal to Nazi-adjacent users. And consider their niche predecessors in enterprises like 4chan or its even scarier descendant Kiwi Farms. Lacking the constraints of market operations at scale, like mass-market users, mass-market advertisers, press attention, and publicly traded stock, these companies had no reason to control the incitement found on their platforms; indeed, it took a large and hence publicly accountable company, Amazon (in its infrastructure rather than platform business), to bring one of them at least temporarily under control, to wit, by kicking Parler off its Amazon Web Services infrastructure following its role in facilitating the organization of January 6, 2021, coup attempt at the US Capitol (Romm and Lerman 2021; Guynn 2021). In a vivid illustration of the effects of social isolation on extremism, at least a handful of the insurrectionists spent some of their time in the Capitol showing off for the presumptively sympathetic audience to be found on Parler (Klein and Kao 2021). According to legal pleadings filed by Amazon in its defense against a suit by Parler over the termination of its hosting service, Amazon had been urging Parler to control the threats of violence against Democratic officials such as Nancy Pelosi at least since November.[36] While Amazon (and Apple, and Google) may certainly be criticized for not putting a stop to Parler well before January 6, there is zero reason to believe that Parler's operators would have put a stop to its incitement on its own initiative *ever*.

This point draws on a larger worry about antitrust: to the extent there's a normative justification for antitrust law in consumer services, it revolves around the idea that a more competitive market will provide a wider variety and better prices of those services to the general public. But where services might be specialized to socially harmful market segments, it might actually be a bad thing to encourage the growth of niche service providers, and a good thing to permit the homogenization associated with big centralized providers.

One might imagine a number of alternative strategies for platform company breakups that could avoid these problems. For example, different services could be disaggregated: Facebook and Instagram could be separated; Amazon's delivery services and its network infrastructure (AWS) could be separated; Google search and YouTube could be separated. There might be some possibility of geographically

don't particularly care (or themselves are interested in purveying or exploiting Nazi-adjacent dispositions), the public prominence of those companies means that their recommendation algorithms are under intense press scrutiny, so that when this has in fact happened (presumably by accident), the news has come out and the companies have tried (however ineffectually) to remedy the problem to avoid being punished by their users and advertisers. For example, Facebook acted swiftly to get rid of algorithmically created anti-Semitic ad targeting categories after ProPublica discovered them (Angwin, Varner, and Tobin 2017).

[36] Defendant Amazon Web Services, Inc.'s Opposition to Parler LLC's Motion for Temporary Restraining Order, p. 4, filed January 12, 2021, in *Parler LLC v. Amazon Web Services, Inc.*, No. 2:21-cv-00031-BJR, United States District Court, Western District of Washington, at Seattle; brief available at https://storage.courtlistener.com/recap/gov.uscourts.wawd.294664/gov.uscourts.wawd.294664.10.0_1.pdf.

centered breakups – such as Amazon logistics across US regions – or along the lines of proxies for geography such as language.[37]

On the social network side, larger companies might be able to achieve socially beneficial economies of scale in content moderation, for example, training machine learning models to identify certain kinds of content on much larger and more diverse datasets. However, in principle, antitrust strategies could be devised to avoid undermining such economies of scale, for example, by breaking up product sides of companies while leaving governance sides (such as "integrity" entities) intact. Voluntary industry cooperation could also ameliorate some of the difficulties of disaggregation in that respect.

On the whole, it seems to me – again, quite tentatively – that while antitrust breakups may be useful to ameliorate other harms that platforms impose on the world – such as invasions of privacy and impeding innovation by blocking market entrants – it is on balance unlikely to be effective as a solution to those problems that fall under the category of "governance" as used in these pages, within the domain of this book's concern, that is, where long-term platform company interests and the interests of the rest of us are aligned.

The question of antitrust, however, fits more closely into the overall project of this book in two respects. First, the democratic reforms which this book defends may also be a component of a program to mitigate the public dangers of antitrust action of the sorts I question above. The core proposal of Chapter 6 is to create democratic institutions of platform governance that interact with, but are independent of, individual companies and operate across them. Creating such institutions might also make it less dangerous to break the companies up because those institutions themselves could embody some of the beneficial effects of scale, such as giving companies pressure to moderate content in accordance with the needs and desires of vast bodies of people (rather than wicked niches) and aggregating both practical knowledge as well as more brute tools like machine learning models or training data. In the world I envision in Chapter 6, governance economies of scale might be capable of being spread across the industry as a whole (depending on how much infrastructure to intervene directly in content moderation and other kinds of rule enforcement could be built around the participatory institutions I describe), and pressure from empowered ordinary people could be directed at the industry as a whole. Under such circumstances, perhaps even small platforms serving niche markets could benefit from overall governance reforms designed for the big companies.

From the other direction, antitrust law could be a *vehicle* for implementing those reforms. As Chapter 4 describes in some more detail, some of the governance

[37] Wu (2018, 83–91) argues that at least some economics of scale in ordinary commercial markets may be illusory. We may well find out after, for example, breaking up Amazon's logistics network into regional sub-units that the smaller companies can still deliver things like affordable and fast delivery. It also might be that logistics is simply a public good which government ought to supply – that is, adequately fund the United States Postal Service.

pathologies of platforms result from poor internal governance, such as the excessive authority of social media company lobbyists over content policy decisions regulating the behavior of the same politicians whose favor they seek. In a recent article, Herb Hovenkamp (2021, 2021–32), one of the United States' most prominent antitrust scholars, suggests using antitrust law to internally reorganize the decision-making structure of platform companies, for example, to empower different groups of stakeholders in order to prevent abuses of power.[38] Hovenkamp's approach is quite similar to my own – he draws, as do I, on Ostrom's governance theory,[39] and offers the compulsory integration of stakeholders in corporate governance structure as a tool to address platform abuses of power similar to – and in some cases overlapping with – the ones I address, such as Amazon's abuse of data to discriminate against third-party sellers.[40] Accordingly, it may ultimately be the case that the reforms offered in this book are less an alternative to antitrust law than a potential product of it.

At this point, I consider the broad approach of a governance development framework adequately defended. The rest of this book turns toward how such a framework would work – the characteristic problems of governance and institutional solutions to those problems shared both by states that struggle to govern and by platforms that struggle to govern (Chapters 3 and 4) and the sorts of concrete institutions that might be built to alleviate them in the platform context (Chapters 5 and 6, and the Conclusion).

[38] It might also be used, for example, as a regulatory lever to force platform companies to permit the third-party creation of certain kinds of "middleware" (D. Keller 2021).

[39] Compare Hovenkamp (2021, 2022); Chapter 3 this volume.

[40] Compare Hovenkamp (2021, 2029); Chapter 4 this volume.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

3

The Problem of Platform Knowledge

Putting a stop to incitement and propaganda aimed at carrying out a genocide is the easiest imaginable moral case for platform governance. The prohibition on genocide is a foundational *jus cogens* norm – a prohibition with as close to universal acceptance as we might want (presumably except by its perpetrators); unlike the more difficult problems discussed briefly in Chapter 2 (such as the Thai lèse-majesté law), there's no basis whatsoever to criticize intervention to prevent genocide as colonial. Moreover, there's a widespread belief among courts, social scientists, human rights scholars, and activists that propaganda like that seen in Myanmar facilitates such genocides (e.g., Wilson 2015; Buerger 2021; Yanagizawa-Drott 2014; Adena et al. 2015; Leader Maynard and Benesch 2016). We can all agree that such propaganda should not be allowed on any platform, and it violates numerous Facebook rules.[1] So what went wrong? The inability of Facebook to deploy local knowledge was the crux of the problem. Facebook lacked reliable ways of finding out about what was happening both on the platform (in the ways to be described below) and in the real world as a result of the on-platform activity.

One incident may be illustrative. In September 2017, "chain messages" were sent over Facebook Messenger to both the Buddhist and the Muslim communities in Myanmar, each warning of impending violence by the other group; according to local organizations, there were "at least three violent incidents" traced to these messages (United Nations Human Rights Council 2018, 341; relying on the account in Phandeeyar et al. 2018). Several civil society groups in Myanmar sent a letter to Mark Zuckerberg relating to that incident which stated that "as far as we know, there are no Burmese speaking Facebook staff to whom Myanmar monitors can directly raise such cases" – and that effectively, Facebook's content moderation system was to wait until civil society organizations sent an English speaker to raise the alarm with the company (Phandeeyar et al. 2018). But the problems ran deeper than local staff: Facebook's software systems weren't set up to

[1] That being said, Facebook saw the need to expand the scope of its rules in order to rectify gaps revealed by the Myanmar genocide (Warofka 2018).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

process the content. Its machine learning systems used to identify things like hate speech couldn't even parse the text encoding in which the incitement was shared (Warofka 2018). While Facebook executives did have *some* warning, as for a number of years beforehand civil society organizations had raised concerns with the company about hate speech in the country (Amnesty International 2022, 51–53), it appears the knowledge didn't lead to action, perhaps because complaints weren't directed to anyone with power, because those making the complaints didn't have the capacity to force the company to pay attention, or because local organizations were unable to convince anyone with authority at the company that genuine physical harms to people were a likely consequence of the propaganda being spread through the platform.

The aftermath of this horror also highlights the fact that the United States is already engaging in colonial platform governance by default. The Gambia appeared in a US court to attempt to subpoena information about the genocidal Myanmar government's communications in order to hold that government accountable at the International Court of Justice. But Facebook opposed the subpoena on the grounds that the Stored Communications Act, 18 U.S.C. § 2702(a), forbade the company from providing the information (Smith 2020; McPherson 2020). In the words of the company, the Stored Communications Act is "a provision of the federal criminal code that protects billions of global internet users from violations of their right to privacy and freedom of expression."[2] (As if the right to privacy and freedom of expression, as interpreted in the United States, applies to a genocidal military on the other side of the world.) Elsewhere in the company's brief, Facebook also makes appeals to a kind of implicit superiority of US law, declaring, for example, that "The Gambia asks this Court to grant it special and unbounded access to account information that no other governmental entity in the world – not even the United States – has available to it in any other proceeding"[3] (as if the level of access of the United States government sets the relevant standard for investigating a genocide), and "Congress's decision to create the CLOUD Act and codify international assistance procedures (e.g., MLAT) reflects its determination of the proper balance between the competing interests at issue in this litigation"[4] (as if the United States Congress gets to decide about the importance of "the competing interests" in a dispute between two totally different countries about a genocide committed by one of them in an international tribunal). Of course, this US-centric rhetoric was necessitated by the fact that The Gambia needed to seek the assistance of a US court to force Facebook to turn over the information, and Facebook evidently saw itself as primarily bound

[2] Facebook's Opposition to Petitioner's Application Pursuant to 28 U.S.C. § 1782 in *In re Application Pursuant to 28 U.S.C. § 1782 of The Republic of The Gambia v. Facebook, Inc.*, case no. 1:20-mc-00036-JEB-DAR (D.D.C., August 4, 2020), p. 1.

[3] Ibid., 9.

[4] Ibid., 9–10.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

by US law (as opposed to, say, that of The Gambia), perhaps because its corporate headquarters and most of its core personnel are located in the United States, which is precisely my point. There is no reason that the people of Myanmar, the people of The Gambia, or the people of the world in general should have to interpret their "privacy and freedom of expression" interests in the ways US law imposes on companies located in its jurisdiction.

One answer to the particular problem faced by Facebook in the Myanmar genocide, had they done so in advance, would have been just to throw money at it: hire more speakers of the local languages as content moderators, task developers to account for minority text encodings. And Facebook surely should have done that.[5] But not even pre-recession Facebook had unlimited resources, and not every knowledge problem can be solved by throwing money at it – one must target one's knowledge-seeking investments, and this chapter is about the organizational background of the conditions under which that targeting can be carried out.

3.1 THE PROBLEM OF KNOWLEDGE: A PERVASIVE CHALLENGE FOR CENTRALIZED GOVERNORS OF DISPERSED POPULATIONS IN CHANGING ENVIRONMENTS

Centralized systems of governance, especially under conditions of diversity, are vexed by the problem of moving knowledge from the peripheries that require regulation to the center from which regulation issues. This problem is the core of leading twentieth-century economist F.A. Hayek's greatest insight: He famously argued that a price system in a free market would be a far more effective regulator of production than any centralized government authority (a la the Soviet Union), because prices capture information about the needs of all buyers and sellers in a market; information that no government agency could ever collect or process. But the problem does not merely occupy the economic domain. Rather, the problem of deploying local knowledge is a key characteristic challenge of states, also highlighted by scholars operating in contexts as diverse as Jane Jacobs's (1992) work on urban planning and James C. Scott's (2008) work on agriculture and central state planning more generally. The commonality among this literature is the insight that the top-down

[5] This is not just a Facebook problem. Whistleblower leaks have identified similar problems at pre-Musk Twitter. In the words of one leaked document ("Current State Assessment," revealed as part of 2022 whistleblower reports and available at https://s3.documentcloud.org/documents/22186781/current_state_assessment.pdf, p. 2), "the lack of diverse backgrounds among employees contributed to gaps in foreign-language and on-the-ground contextual capabilities, hindering Twitter's ability to execute its mission and remove harmful content worldwide. Teams in priority growth markets either do not exist, or are not sufficiently staffed or resourced." As I wrap up the manuscript of this book, it has come out that the major social media companies are still failing at what seems to me to be a much easier case. However badly they're doing at controlling Russian state propaganda about its aggression against Ukraine in English, they're doing even worse in other languages – even in languages spoken widely in the United States like Spanish (Myers and Frenkel 2022).

viewpoint of a state may be unable to effectively carry out its policies because of its inability to determine the consequences of those policies (or even the needs to be met) in particular local contexts which may be different from the epistemic environment of the policymakers.

I shall suggest that those insights apply just as well to private governance – actually more so. First, however, we should delve a little bit more into the intellectual background to get a fuller sense of the commonalities among the settings of and proposed solutions to the problem of knowledge.

Jane Jacobs's *The Death and Life of Great American Cities* turned the profession of urban planning on its head in 1961.[6] In a reaction against centralized models of utopian urban planning associated with architects and planners of the likes of Olmstead, Moses, and Le Corbusier, Jacobs understood urban planning (and larger economic processes – see discussion in Desrochers and Hospers 2007) as fundamentally organic, bottom-up, improvisational, and adaptive. Centralized planning – the gigantic housing project, the "slum clearance" – in such a context could be actively value-destroying, because existing urban forms of organization, even if they looked messy, represented successful adaptations to the challenges distinctively faced by the people of a place in the context of their own social, economic, and physical lives – in Jacobs's (2016, 75) words, they incorporated a kind of "weird wisdom" derived from the activities of the people occupying that space.

For Hayek, it would be easy for a policymaker to be tempted by Soviet-style central planning. For the central planner, it might seem sensible to make a decision like "we need this many tires, this many gallons of oil," and to governmentally control production in order to generate those goods. Our planner thus avoids all the complexities of free markets, such as inequalities of wealth (and hence of access to human needs), wasteful production of unnecessary luxuries, redundancies in suppliers, and so forth. But in identifying the trap into which our planner has fallen, Hayek (1945, 524–26) too focused on dispersed knowledge and on adaptation: It turns out that it's actually impossible for central planners to know things like how many tires are needed, because that's just the aggregate of the needs of everyone else, which are hard to discover; moreover, even if the central planner could know how many tires were needed at a specific point, it would be impossible to adapt to changes in local conditions which might shift where and how many tires are needed in an instant. However, free markets – and in particular the price system of such markets – could effectively aggregate that information, because people respond to local changes in needs by changing their willingness to pay. If tires are particularly needed in a specific community, they will out-bid other communities for those tires, and thereby communicate their need to producers – with accompanying incentives to take that information into account – without any of the participants needing to know anything other than the change in the prevailing price.

[6] See discussion in Campanella (2011).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

Observe how Jacobs and Hayek share an emphasis on a key problem that applies in the platform context with some salience: *adapting pre-existing plans to change, where information about that change is spread out across the governed territory*. Moreover, the key method of such adaptation is to give those with local knowledge about conditions and needs some direct *control* over outcomes. Thus, a planner who gets the answer to an immediate problem right and makes commitments – whether those commitments are the resources to be allocated to particular production and distribution networks or the buildings to be built – cannot adapt those commitments to exogenous shocks unless people with both the knowledge and the incentives to respond to those shocks are capable of exercising some control over them. Both Hayekian price systems – which encourage producers to change pre-existing resource commitments in pursuit of profit – and the organic development of urban landscapes – which permit relatively fast-paced and small-scale adaptation – can serve those adaptive functions.

Some decades later, James C. Scott (2008) integrated the – by then widely accepted – insights of both Jacobs and Hayek into a broader analysis of the failures of a style of centralized government intervention that he called "high modernism." In a way, on Scott's account, centralized authorities recognize that they have a problem with learning the information necessary for their projects to succeed – and their response is to attempt to impose "legibility" on the world. The idea of legibility effectively highlights a core idea shared by Hayek and Jacobs, namely that the distinctive informational problems with which they wrestle are the product of complexity: Cities, like economies (and, I submit, like platforms) feature dense interactions among numerous and diverse actors, exponentially increasing the informational challenges associated with centralized planning. Thus, for Scott, the response of state bureaucracies is to attempt to reduce the dimensionality of the problems with which planners are presented: to summarize them only in terms of the particular issues with which they are concerned as matters of policy and then, in effect, attempt to optimize along those lines (compare Reich, Sahami, and Weinstein 2021 on pathologies resulting from narrowly focused technology company optimization). But such a myopic approach tends to lead to disaster, as the needs and values as well as the local knowledge and adaptive capacity of those whose lives are thereby reorganized tend to be disregarded. Thus, Scott describes the failure of a wide variety of high-modernist planning projects, ranging from Soviet and colonial agricultural "reforms" to (in passing) Taylorist factory management to the creation of planned cities like Brasilia.

In the abstract, we might summarize the core lesson of Jacobs, Hayek, and Scott as a pattern for the solution of these kinds of knowledge problems. When important knowledge is in the periphery and cannot be effectively directly moved to the heads of central decision makers, individuals must first be given an incentive to use their knowledge – that is, doing so must produce useful outcomes for them as individuals. Those uses of knowledge must also, in the aggregate, be capable of correctly influencing aggregate outcomes. Hayek's account of prices, thus, accounts for both the individual incentives – people need to use their knowledge of their own

needs and resources in offering and demanding goods in markets – and the aggregate influence – the overall prices (under the right conditions) successfully reflect the conjunction of needs and resources of all in the market. Similarly, the existing urban structures in Jacobs's weird wisdom represent the successful use of individuals' knowledge to solve individual problems but are (quite literally) built into urban structures that continue to be functional and to interact productively with similar uses of knowledge from other individuals.

Many of the problems of platform governance can be understood as center-periphery knowledge problems of the sort considered by Jacobs, Hayek, and Scott.[7] Indeed, based on what we know about the class of problems, we ought to have expected them to be the characteristic challenge of platform governance. For such problems are rooted in scale and diversity: As the number of people the governors have to regulate grows and as the social distance between the governors and the governed grows, it becomes more difficult for those doing the regulating to have any idea either about the needs or the activities of those regulated. Add in the capacity of platforms to create network configurations both dense and dispersed in ways previously unavailable to human sociality, and the emergent patterns of behavior that we've all watched result from that capacity over the last few years – as well as the consequent high rate of innovation, novelty, and chance – and it should be unsurprising that platforms have had persistent problems with learning about user behavior and interests as well as adapting to changing behavioral environments.

Thus, consider as a salient example the various missteps of over-governance on social media – circumstances where social media companies have misinterpreted subcultural communication or failed to attend to the local context, and, in doing so, created scandals over their removal of socially beneficial content. Just on Facebook alone, for example, there have been famous scandals involving the removal of famous Vietnam War Photos (Scott and Isaac 2016), requiring transgender people to use their birth (dead) names (Holpuch 2015; see discussion in Haimson and Hoffmann 2016), removing as hate speech content of LGBT users who are reclaiming terms like "dyke" and "tranny" (Lux 2017), and treating topless photos of (and posted by) indigenous women in traditional settings as forms of obscenity (Alexander 2016). Nor are other companies free from similar problems; for example, researchers found that Google-developed artificial intelligence to detect online "toxicity" disproportionately identified the speech of drag queens as toxic (Dias Oliva, Antonialli, and Gomes 2021). Even as recently as 2021, these problems persist, even in Burmese (the one language you'd think Facebook would

[7] Frank Pasquale (2018) has a particularly interesting take on platform knowledge. He begins with the notion that platforms are in some sense a *solution* to the problem of knowledge, insofar as their immense amount of data and machine learning capacity permits them to organize far more activity than the state when Jacobs, Hayek, or even Scott were writing. But this turns out to be illusory, for platforms too struggle to manage such a vast scope of activity, and they end up becoming "absentee landlords" when they cannot manage their gigantic domains (or choose not to care because neglect is profitable).

try to get right after the genocide): When Burmese speakers advocated borrowing financial strategies from Hong Kong's resistance to the Chinese Communist Party to resist Myanmar's authoritarian government, Facebook misinterpreted it as hate speech against Chinese people, necessitating a reversal by the Oversight Board.[8]

Moreover, as James C. Scott would predict, efforts to address the problem of knowledge by rendering human behavior on platforms legible can distort governance choices. For example, Siapera and Viejo-Otero (2021) describe how Facebook's efforts to address hate speech have manifested as the kind of "colorblind" (and gender-blind, disability-blind, colonialism-blind, etc.) rules critical race theorists have aptly criticized in US law for failing to take account of underlying social hierarchy in interpreting hierarchy-laden concepts like racism (e.g., Gotanda 1991). Although Siapera and Viejo-Otero read this colorblindness through an ideological lens, they also discuss what seems to me to be a more primary cause, if not of the adoption of colorblind policies then certainly of their resilience against external critique: The difficulty of identifying the relevant relationship between social hierarchies and language in different societies, that is, a problem of legibility. This is illustrated by an extended quote from Mark Zuckerberg, which leads Siapera and Viejo-Otero's article and which attributes Facebook's colorblindness to the problem of dealing with vast diversity in social meaning at scale:

> [W]e don't think that we should be in the business of assessing which group has been disadvantaged or oppressed, if for no other reason than that it can vary very differently from country to country [...] It's just that there's one thing to try to have policies that are principled. It's another to execute this consistently with a low error rate, when you have 100 hundred billion pieces of content through our systems every day, and tens of thousands of people around the world executing this in more than 150 different languages, and a lot of different countries that have different traditions.[9]

3.2 THE DEMOCRATIC, DECENTRALIZED SOLUTION TO THE KNOWLEDGE PROBLEM

Broadly speaking, we can think of Hayek and Jacobs as offering solutions to problems of knowledge rooted in individual agency – individual decisions of buying and selling in a market which aggregates those decisions for Hayek; individual decisions in land use aggregated by the fact that those decisions are also influenced by the nearby individual decisions of other land-users for Jacobs. But these are modes of knowledge aggregation that most plausibly occupy what we might think of as *productive* contexts, as opposed to *regulatory* contexts – that is, circumstances when

[8] Content Moderation Oversight Board, decision in case FB-ZWQUPZLZ, https://oversightboard.com/decision/FB-ZWQUPZLZ/.

[9] Siapera and Viejo-Otero (2021), 113, quoting leaked statement from Mark Zuckerberg.

people are building or consuming things, and they have strong incentives to make the things they build or consume compatible with one another. It is less obvious how such patterns might be adapted to a purely *regulatory* context like platform governance. In conventional polities, for example, we do not let people evolve bottom-up rules of criminal law from their individual choices.[10] Instead, a different structure of aggregating individual knowledge is needed, and the standard approach is to connect democratic political decisions to collective learning. Dewey (1927, 206–8), for example, argued against expert government on the basis that democratic processes permit ordinary people to communicate their knowledge and needs upstream.

The abstract logic of this argument is tempting: Democratic institutions give people an incentive to contribute their knowledge – their votes may produce outcomes that are to their benefit – and translate those uses of knowledge into collective outcomes. Democracies start to sound much like Hayekian markets. But – in part because this idea comes so naturally to those of us steeped in the ideological environment of contemporary western liberal democracy – the democratic solution has been vexed by a number of important challenges to both ends of the solution pattern. On the collective outcomes end, a key challenge has come from social choice theory, which captures the idea that it's surprisingly difficult to translate individual preferences (hopefully reflecting knowledge) into aggregate outcomes (e.g., Riker 1982). On the actual knowledge end, it's widely recognized that democratic institutions tend to give people votes on things about which they don't necessarily have particularly meaningful knowledge, and, moreover, may give individuals insufficient incentive (due to the low probability that their votes may deliver any desired outcome) to acquire or use relevant knowledge (e.g., Achen and Bartels 2016; Brennan 2016).

And yet there is also an important literature on the capacity of democratic institutions – if the incentives are right (e.g., if there are shared preferences as to common goals) – to structure the effective aggregation of knowledge in ways that at least potentially avoid these problems. Particularly important is Ober's (2008) account of how the network structure of Cleisthenes' reforms in sixth-century Athens helped the *demos* effectively leverage dispersed knowledge (creating, I daresay, the very first networked leviathan). On Ober's account, Cleisthenes' system combined strong-tie network structures in local demes (essentially, neighborhoods/villages) which permitted people to develop dense social knowledge about the knowledge, skills, and character of their neighbors, with artificial forms of higher-level social structure that permitted the formation of weak network ties between demes. These included "tribes," which were administrative and political units which were created to combine demes from coastal, inland, and urban regions (hence building geographically dispersed network ties), and the Council of 500, the main executive body of the city, which was made up of randomly selected representatives from each tribe.

[10] I confess that we might tell a story about American criminal law's emergence from English common law which interprets them that way.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

Ober teaches us that the network structure of democratic decision-making matters. In the abstract, he offers a case for a structure that promises people influence (either locally or by integrating them into centralized decision-making mechanisms in such a way as to give them a visible impact on decisions at least some of the time, as by sortition-based councils), thus giving them an incentive to put their knowledge to use; as well as broader and socially novel network connections, to transmit knowledge among individuals and to socially distant decision makers.

Note how Ober's (2008, chapter 4) account of Athenian knowledge aggregation mechanisms describes key functions to handle the problem of aggregating, centralizing, and retaining diverse knowledge. The Council of 500 aggregated widely dispersed knowledge from all sectors of society, and operated essentially continuously while giving members incentives to build novel network interconnections with unfamiliar persons. Individuals only had short terms on the council and were likely to receive reputational rewards for successful leadership. They had strong incentives to interact with people who would otherwise be socially distant from them in order to succeed in their Council roles. After their terms were up, they were also likely to reap economic benefits from the distant connections that council service made available. Moreover, because of the density of council service within the community (due to short terms and sortition), that knowledge could be further dispersed as well as transmitted to new members.

What might all of this mean in practice in the modern context? Specific proposals will be left to Chapter 6, but Athens-like structural innovations seem tailor-made for handling platform problems involving dispersed knowledge. Imagine how the Myanmar disaster could have gone differently if a local (and socially dense, in a network sense) group of Burmese citizens and activists capable of swiftly and credibly identifying the presence of military propaganda on Facebook was directly incorporated into company decision-making processes via some similar kind of council and had already built relationships, via those processes, both with corporate personnel and with similar local groups elsewhere? Suppose that Facebook's internal bureaucracy was already accountable to such a council, which was diverse but included ordinary Burmese people, who regularly met to consider updates to platform rules and enforcement priorities? Such a council could have escalated the risk of genocide to the company earlier – and would have had an incentive to do so, if its word actually counted in company decision-making (i.e., if the participation of ordinary people was efficacious).[11] Or, perhaps, it could have escalated the

[11] Had Facebook been alerted of the danger, it's likely that the company could have at least done something relatively quickly, such as a patch to their content moderation tools to handle the text encoding in which the hate speech was rendered at a minimum. Some suggestion that there was low-hanging fruit available to at least mitigate the risk comes from the fact that after the scandal, the company supplemented its automated content moderation and managed to begin to do a substantially better job in Myanmar (Gorwa, Binns, and Katzenbach 2020, 2).

problem to international human rights agencies. Put a different way, it could have leveraged local knowledge to more readily identify community needs.[12]

Note how this is distinct from the task of merely hiring more content moderators. In the platform economy, as it currently stands, content moderators are substantially disconnected from actual company decision-making processes: As Sarah Roberts (2019) describes, such moderators are typically hired by contractors, physically isolated from company decision makers, and managed in a high-rate task-oriented fashion that gives them little opportunity either to deliberate about the problems they see or communicate with decision makers about priorities.

Another way of putting the problem is in Scott's terms. In the first instance, what matters to Facebook is the match between content on the platform and the rules they've written down; the job of content moderators is to make that content legible in those terms. Accordingly, the sorts of knowledge which content moderators may leverage are chosen in advance by the company: They may use their knowledge of the local language and idioms to identify whether some particular platform activity violates a closed set of rules. They are not asked to use their knowledge about the overall conditions and needs of the communities of which they are members. Even moderators who have worked on the main company campuses as contractors have reported that their input is not welcome (Roberts 2019, 97).

That knowledge is, from the perspective of platform governance, simply wasted – but it need not be wasted and could be deployed with the creation of democratic decision-making processes accountable to people in those communities. Indeed, this could include the content moderators themselves along with ordinary users, to the extent they could win or be granted (e.g., by states) labor rights entitling them to some degree of participatory decision-making in their workplaces. More empowered content moderators with a real voice in their workplaces and real career paths would be in better positions to communicate knowledge that they have to company policymakers.[13]

Chapter 6 will propose the creation of randomly selected grassroots governance organizations across the world as the primary method to integrate peripheral knowledge into platform companies. Hélène Landemore, another key democratic theorist of knowledge, has argued (Landemore 2013, 2020), based on the analytic work of scholars of diversity and democracy such as Hong and Page, that sortition

[12] Compare Anderson's (2006, 17–21) discussion of Bina Agarwal's (2000; 2001) work on forestry decision-making in South Asia, which offers a Deweyian interpretation of Agarwal's research on the way that the exclusion of women from many local groups has led to failed decision-making because that exclusion also sacrificed knowledge about things like how to enforce resource management rules and how much wood may be safely gathered.

[13] Organizational sociologists have noted that personnel moving between roles is a key mechanism of organizational and intraorganizational learning and may also contribute to "redundancy in mechanisms for sensing and responding to change" Westley (1995) 416–417. Making content moderation less of a dead-end job and more of a job that could rotate into company decision-making roles as well as jobs with regulators and the like has additional potential to more fully leverage local knowledge in this sense.

is epistemically superior to other methods of democratic selection in virtue of its capacity to generate diverse pools of decision makers. (Classical Athens, not incidentally, is famous for its use of sortition to fill decision-making bodies.)

Sortition is likely to be a key selection process for any democratic form of platform governance for several additional reasons in addition to its positive effects on the ability of such processes to leverage knowledge from persons socially distant from platform employees. First, in the context of extreme global institutional diversity, sortition seems likely to be much easier to administer than other selection methods, as it would not require adaptation to specific versions of more active political selection (such as campaigning and elections) with which users may be disparately familiar. Relatedly, sortition at least partly routes around institutional corruption and subversion – to the extent any other method of selection requires active administration of things like voting rules by either companies or governments, sortition reduces the risk that such administrative processes can be used to put a thumb on the scale.

That being said, it is unlikely that entirely sortition-based democratic processes will be sufficient. There is also strong reason to guarantee (and not just probabilistically) representation from a wide array of discrete social groups in platform decision-making. Moreover, sortition would be most practicable to administer from among platform users, but nonusers and groups predominantly composed of nonusers are also the victims of platform externalities and are likely to have helpful knowledge to contribute about it (as well as moral claims to participation in its governance). Nonetheless, the advantages of sortition are sufficiently compelling that it will certainly be a key component of any effective democratic platform governance system.

Another knowledge-related advantage of the creation of broad-based participatory democratic institutions flows in the other direction, that is, from platforms to the public. As Cohen (2019, 181) has aptly argued, regulatory strategies for things like the abuse of platforms that rely on transparency are unlikely to be successful, because under conditions of "infoglut," the challenge is not accessing information but interpreting it. (Have any of the readers of this book ever looked at the ad transparency database Facebook rolled out after the 2016 disasters?) But one solution to infoglut is specialization, that is, to delegate to a part of the global public at large primary responsibility to process such information. Traditionally, such roles are held by researchers and journalists, but both are also subject to infoglut,[14] as well as to other institutional pressures as well as a lack of direct access to relevant decision makers. Participatory democratic platform governance institutions could in principle be given such direct access, for example, to regulatory officials of the states in

[14] Go ahead, just ask me about how easy it is to turn in a book manuscript within a year of the deadline written in the contract about something as timely as platform governance, about which there is a constant flood of new scholarly publications as well as evolving facts on the ground. In unrelated news, my most sincere apologies to my long-suffering editor at Cambridge and grant administrator at the Knight Foundation.

which platforms operate, and may benefit from sufficient specialization over a short term to be able to focus on and process information coming from platforms, avoiding the problem of infoglut. (We might compare this to the specialized magistrates in Athens who took time off from their ordinary lives to manage particular areas of Athenian government.)

In addition to giving people an incentive to contribute their knowledge,[15] participatory governance can also bring together knowledge and legitimacy. The sociological concept of the legitimacy of a system of governance refers to the notion that people will actually accept governance arrangements and hence will comply with the outcomes of those arrangements voluntarily rather than by force (Tyler 2006). One common proposition in the literature on sociological legitimacy is that voice or participation is a source of legitimacy in this sense (Tyler 2009, 319).[16] There is also evidence that people given an opportunity to participate in jury-like resolutions of social media content moderation cases perceive those processes as more legitimate, although not necessarily more effective (Fan and Zhang 2020).

We might, however, helpfully disaggregate the relationship between participation/voice and legitimacy into two plausible causal directions. First is the conventional supposition that to the extent people are given an opportunity to participate in rulemaking and application/enforcement processes, they will perceive the decisions made under those rules to be more legitimate, and hence be more likely to comply with them. The second, perhaps more interesting, idea reverses the causal direction and observes that participating in collective decision-making is costly (as any reader who has ever been in a faculty hiring meeting can attest), and hence that recruiting people to participate requires giving them some reason to think that the ultimate decisions will be something they can accept – for example, by giving some guarantee that their participation will be effective at preserving their core interests, that is, providing avenues for what we might (mildly tendentiously) call *real* or *meaningful* participation. This reverses the causal relationship in the sense that it derives participation from legitimacy – it suggests that the creation of opportunities for *meaningful* participation, opportunities that respect the interests and status of potential participants, might help give those thus respected a reason to participate.

Both of these causal directions are at play in the claim of this chapter that dispersed and distributed governance, done right, has the capacity to improve effectiveness in two ways: by recruiting the knowledge of participants and by recruiting their voluntary compliance (and hence reducing monitoring and enforcement costs). But the *meaningful* participation proviso has other important benefits, for sometimes people who only enjoy nonmeaningful participation are simply ignored

[15] Cf. Fung and Wright's (2003, 25) related concept of "empowered participatory governance."

[16] There is some question about the empirical robustness of this proposition (Hibbing and Theiss-Morse 2008). Still, there is some evidence from the new-governance-style networked governance arrangements on which this volume partly draws that greater degrees of voice are associated with greater degrees of perceived effectiveness (Klijn and Edelenbos 2013).

when they try to contribute their knowledge. This too happened in Myanmar with Facebook: Civil society organizations tried to tell Facebook what was going on but were ignored (Amnesty International 2022, 51–53). Had those organizations or ordinary Burmese people the power to directly intervene on content moderation decisions in the country, they likely could have gotten more serious attention from more powerful company personnel, if only because appeasing them would become more necessary in order to meet more of the company's overall goals.

3.3 PARTICIPATORY GOVERNANCE FACILITATES LEGIBILITY

It will be useful, to fill out the character of the knowledge that ordinary people as well as presently disempowered workers might introduce in some more detail, to disaggregate the problem of local knowledge into distinct problems of *observability* and *legibility*. This is not meant to be a robust typology but rather a pair of ad hoc constructs (partly derived from Scott) to help us distinguish the ways that governments might be challenged to know about behavior and interests and the ways that platforms face similar challenges.

Legibility as a concept has been used prominently in studies of the institutional design of state authority (Scott 2008) as well as urban design (e.g., Lynch 1960). For present purposes, we can think of legibility as referring to the capacity of an observer, having observed some behavior, to classify it in some pre-existing conceptual scheme. The concept of legibility is important in the nongovernance platform context, because, as Julie Cohen (2019, 37–40) explains, providing legibility of persons to advertisers is a key economic function of platforms. In the present work, however, I focus on the problem of platforms acquiring legibility over user behavior.

What might it mean for a platform to "know" what some behavior "is?" Typically, platforms have a high capacity to observe conduct on their platforms. After all, a key distinction between internet platform companies and every other form of governance is that every interaction within the governance scope of platform companies is hosted on infrastructure in the control of the platform; with the exception of end-to-end encrypted communications (most importantly, WhatsApp) and certain kinds of decentralized technologies such as the Mastodon federation,[17] the companies under discussion here have the power to inspect the technical contents – the ones and zeroes – of any byte flowing through their network assets – any financial transaction, any Instagram image is visible in that limited sense.[18] This is in contrast

[17] As a decentralized network of platforms, the Mastodon federation has the potential to upset many of the suppositions on which the platform economy runs. However, the technology is only coming into relatively widespread use after the Musk takeover, and hence has not been observed sufficiently long for me to make any confident statements about it; moreover, it may not support for-profit financial models.

[18] Blockchain and other kinds of decentralization technologies may change this; however, I lack the expertise – or, quite honestly, the interest – to consider blockchain; it is also very difficult at best to come to any plausible guesses as to the likely impact of blockchain on all the business areas currently

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

to governments, which must expend resources on surveillance to observe any particular thing happening in their domains of governance.

For a concrete example: As I type the words of this chapter, chances are that the government has no clue that I am doing so. As far as I know, there are no government cameras in my office; no prosecutor or intelligence agency has gotten a warrant to search the contents of my computer. Maybe the NSA is hoovering up the data that my machine sends over the network, but it doesn't directly control the software on my computer and I'm not remotely important enough to warrant the attention of serious government hackers, codebreakers, exploiters of zero-days, and so forth, so, as long as the various encryption algorithms that my computer automatically uses to send my writing around hold, I can feel pretty confident that the government won't "know" this paragraph exists, in the extended sense of "know" that means "some government official can look at it without getting a warrant to intrude on someone else's property rights by force" at least until after the book is published.

By contrast, at least one platform company can look at it at will. I store my working drafts in markdown files in private repositories on GitHub, which is owned by Microsoft. For all practical purposes, Microsoft can observe this text shortly after I write it. And, in all likelihood, were I to include some content in this draft that violated an important legal obligation either to Microsoft or of Microsoft – for example, if I uploaded the Windows source code to the repository for this book, or child pornography – automated systems would probably detect that and take the content down and/or alert Microsoft personnel long before the government could find out.[19]

However, they may lack the capacity to *interpret* content on their platforms, that is, either to translate bits into communication (as with the task of text and photo identification), or to discern the meaning of communication (as with linguistic translation as well as subcultural message interpretation). This is the problem that

encompassed by the platform economy. The technology has been around for over a decade now, and doesn't yet seem to have had much impact beyond creating a bunch of extraordinarily energy-hungry speculative bubbles. (Also, I think, albeit with very low confidence, that blockchain actually would increase overall observability of some activities to the extent it replicates records of those activities in a distributed database?) As for end-to-end encryption, to my mind this is a feature of platforms that they themselves control – Meta doesn't have to offer end-to-end encryption on WhatsApp; its executives choose to do so. In principle, if they have a sufficient interest in controlling conduct carried out through that platform the company could cease encrypting it. Alternatively, some kinds of governance could in principle be pushed to the application layer – for example, by building certain machine learning classifiers for content like hate speech directly into client applications (which see cleartext) rather than the network, subject to available computing resources on client devices (which may be a hard constraint in some particularly sensitive markets for services like WhatsApp, such as India).

[19] That's speculation, but it's pretty plausible speculation: I mentioned child pornography as one of my examples because Microsoft has actually invented one of the leading technologies to detect and block known child pornography images; given the extremely harmful and criminal nature of such content, it would be surprising indeed if they didn't have the sense to use it on Github. See Microsoft, PhotoDNA, www.microsoft.com/en-us/photodna (visited April 3, 2021). (Microsoft describes some of its own use of the technology on its services in Langston 2018.)

I call "legibility" and it should by now be obvious that I do so because it's useful to swipe some ideas from James C. Scott. Interpretation depends on context: External knowledge can make a given observation more or less legible.[20] The interpretation of idioms is one example: Speakers of British and American English typically have different knowledge about the meaning of sentences including local slang; one person's homophobic slur is another's cigarette. But noncommunicative intentions can also be the subject of interpretation. For example, Facebook engages in efforts to police "coordinated inauthentic behavior" where different accounts work together to deceive users (Gleicher 2018). Such coordination is often facilitated by offline interactions which Facebook cannot directly observe, with the quintessential example, of course, being the "Internet Research Agency," a Russian intelligence operation to spread political disinformation across social media. Interpreting a given piece of content as genuine political advocacy, as opposed to coordinated inauthentic behavior, depends on extrinsic knowledge about, for example, the identity of the speaker and their other social affiliations.

Intuitively, and without delving into the philosophy of language, we can fairly confidently assert that legibility comes in layers, where understanding of a layer depends on the external knowledge one brings to an interpretation. A stream of bits passing through some router might be legible as text rather than an image if one can interpret the metadata attached to its network packets. With a little more knowledge (about which text encoding is used), the text can be deciphered into letters. With still more knowledge, those letters can be assembled into words in a known language. The meaning of those words further requires social context: What a given set of words in a particular language means to speakers and listeners in one social milieu may be completely different from those in another. The obvious concrete example of that last stage with which most readers will be familiar is the deep complexity surrounding the use of certain words which have traditionally been race or gender-based slurs: Used from a member of an advantaged group to a member of a historically oppressed group, such slurs can be among the vilest of hate speech; the same words can also be reclaimed and used within historically oppressed groups as a signal of affinity and solidarity.

For an observer, legibility is a classification exercise relative to a particular task.[21] Seeing a stream of bytes passing through a network, an observer interested in knowing whether it is pornography and one interested in knowing whether

[20] For example, PhotoDNA renders raw bits of image data more legible as child pornography by matching them against context, where that context is the universe of images previously identified by human observers.

[21] This is also true of the other sort of legibility in which platforms specialize, namely the legibility of users to advertisers. As Cohen (2019, 71) notes, the purpose of that kind of platform legibility "is not understanding but rather predictability in pursuit of profit"; in other words, an advertiser does not care whether a platform's profile of an individual is accurate, it cares whether a platform's model can accurately predict which users are likely to buy their products.

it is political dissent may enjoy different levels of legibility, depending on their capacity to understand the ways in which the sender and the receivers of the bytes express their sexuality or their politics. This is Scott's (2008, 22–23, 80) point in deploying the concept: Raw unmediated data from the world isn't usable in any project of governance; rather, that data must be slotted into administratively meaningful categories; to translate from some-stuff-happening-out-there to "this is political dissent," "this is porn" is to make it legible. For that reason, the interpretation of our social world is typically a goal-directed activity. In fact, depending on the interests of an observer, the same image might be pornography *and* activism – for example, a sex workers' rights activist might make activist pornography, which may be interpreted just as pornography (and hence a fit subject for regulation) by a platform employee tasked to enforce rules protecting children from sexual content, and just as activism by a platform employee or user interested in promoting debate on sexual freedom.

A point that I have not made explicitly but which should be apparent from the last two paragraphs is that many of the most important kinds of legibility at play in both governments and platforms depend on inferences about the cognitions that actors, speakers, and listeners have in their heads. This is not to take a position on some kind of deep question of linguistic meaning – I don't care whether or not the author is dead – but simply because the *practical purposes* of those who try to interpret communication often involve drawing inferences about the intentions of speakers and actors in addition to things like the effects of such speech on listeners and observers. Thus, in order to figure out whether some transaction carried out on a platform is actually an illegal arms or drugs sale, we need to know whether the messages that are sent over that platform constitute a shared intention to cause one participant to send some weapons or drugs to the other. In many cases, the contextual knowledge required bears on the pre-existing set of assumptions and understandings shared by the users of a platform which allows them to form these shared intentions. This is why the hate speech problem is so difficult: Hate speech is deeply contextual; the n-word (and related spellings and pronunciations) has very different meanings depending on a wide variety of facts about who is addressing whom with the term, where even the most obviously relevant facts (i.e., the racial identities of the people involved) are themselves complex and context dependent (Gowder 2015).

Because of the vast cultural diversity on platforms, the contextual knowledge necessary to attain high levels of legibility is likely to be particularly difficult for platform companies relative even to governments. Returning to the example of the private GitHub repository that contains the draft of this book: Microsoft is likely to be able to identify the contents of this repository as English text – it's in a standard UTF-8 text encoding; machine language recognition is easily sophisticated enough to identify the English language from a sample of hundreds of thousands of words – and in view of the subject matter, it's likely that most any Microsoft employee who looks

at it will have a general idea of the content. That would not likely be the case if this were instead a book in, say, Tamil, and if the subject matter were, say, the influence of Immanuel Kant on evidentiary approaches in eighteenth-century Bulgaria. However, even with respect to this book, Microsoft would be obliged to spend some resources to reach a very advanced level of legibility – the book makes references to scholarly conversations which some, but not all, Microsoft employees would have the background to interpret without doing additional work.

Wilson and Land (2021, 1060–69) describe the problem of platform regulation of hate speech in similar terms. On their account, a key issue is that such regulation lacks sufficient "context" – where that context derives from an understanding of the underlying community. Even if a company can identify something like hate speech in the abstract – saying nasty things about an ethnic group, for example – social context is required to understand the difference between, say, a group that is socially subordinated complaining about that subordination and incitement by a dominant group to violence. Moreover, the practical impact of that speech, as they note, may differ according to local context about which moderators may be unaware. In their words:

> A content moderator located in a distant country may not be aware of widespread election unrest, outbursts of communal violence, or a pattern of violence against sexual minorities in a locale. Since each moderator only sees a small sliver of the total range of expression about a topic or person, campaigns of systemic harassment are harder to identify.[22]

Even though, as Scott detailed at length, legibility is a core problem for governments, nonetheless, for most kinds of communication, the greater linguistic and cultural proximity between governments and their people is likely to facilitate legibility, relative to platforms.[23]

Now return to Facebook in Myanmar. Unlike a local government, Facebook would not necessarily – and at the critical points when it should have denied its services to the perpetrators of genocide did not – have personnel with linguistic and cultural competence to interpret communications over its network assets in Myanmar. Accordingly, where a local government might be able, could it observe certain messages transmitted over Facebook assets, to understand what was being said, Facebook could observe such messages at will but not interpret them.

Of course, in the case of the Myanmar genocide, the government couldn't have helped, as it was a perpetrator. But if we imagine a counterfactual nongenocidal

[22] Wilson and Land (2021, 1067).

[23] The classic example of the gains to legibility from cultural proximity comes, of course, from the famous Navajo code-talkers of World War II: because the United States had a much easier time than Germany or Japan in finding and hiring people from the Navajo Nation, the cost to make communications in code in Navajo legible was much, much lower for Americans. Apparently a key advantage of the code-talker system was the distance between the sounds produced in Japanese language and those produced in Navajo, which stymied even consistent transcription of the coded communications (Huffman 2000).

government trying to address private speech that threatened ethnic violence against the Rohingya people, the contrast between that government and the platform couldn't be clearer: Facebook had the capacity to observe every message sent through its servers but couldn't actually figure out what it meant without third party assistance or costly (and all-too-late) investments in native speakers of the language who understood the culture (and technical investments in parsing the text encoding called Zawgyi). The (counterfactually nongenocidal) government was filled with native speakers who understood the culture, and so in principle could understand most messages transmitted over the platform, but had no practical way to observe those messages in the first place.

However, one of the key lessons from scholars like Scott and Jacobs is that even achieving legibility in the limited sense of knowing at a sufficient level of detail what some language means is insufficient. The discussion thus far is still within the realm of hiring more content moderators, at greater or lesser degrees of individual empowerment – to interpret individual pieces of content in light, that is, of the ability to use greater or lesser degrees of local knowledge about the meaning of that content. But some sorts of legibility are in some sense only accessible in the context of a design that is itself derived from local knowledge rather than centralized command. This is part of the point of the theoretical work on democratic knowledge I described a few paragraphs ago: Another way of reading Ober's insight is that democracy creates feedback loops that can make use of local knowledge. A centralized decision which, for example, creates rules exploitable by a genocidal government cannot be rectified just by taking down more and more hate speech; rather, that decision itself needs to be subject to *feedback* based on knowledge about things like the actual local threat of genocide, which can then be routed back into the design of rules and enforcement systems sensitive to the local context.

We can borrow from one more participatory context to learn about the usefulness of participation for legibility, to wit, the common law. In American law, contextual knowledge (or at least judgment) is often specifically incorporated into the criteria for triggering legal consequences. For example, in articulating the test for whether some speech is outside of the protections of the First Amendment on grounds of obscenity, the Supreme Court has declared that an obscene work must be one as to which "the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest."[24] This reference to "contemporary community standards" directly invokes the contextual knowledge of some (not terribly well-specified) community external to the legal system whose standards are to be used. Consider also the appeal to the "reasonable person" in setting the standard of care in negligence law.[25] In general, the common law frequently appeals to community understanding captured in various reasonableness doctrines.

[24] *Miller v. California*, 413 U.S. 15, 24 (1973).
[25] See discussion in Gilles (2001).

It's no coincidence, I think, that common law systems such as the United States and England also feature the substantial use of juries. There's a kind of inherent connection between those two systems: Part of the underlying theory of the common law has been that it is genuinely common, that is, that it derives from the experience and values of the community in organizing their lives.[26] And sole elite judges are naturally less competent in making that judgment than a diverse group of ordinary people from the community, in a deliberative setting in which they can draw on and reconcile their own experiences of community norms.

There is also a close connection between the form of the jury and the resistance to the oppression of minorities, as the demand to be included in a jury can amount to a demand to have one's own knowledge and judgment included in the process of applying the rules used against one. The famous Ten-Point Program of the Black Panthers expressed the point most vividly: because Black people were excluded from juries, "We have been, and are being, tried by all-white juries that have no understanding of the 'average reasoning man' of the black community" (Bloom and Martin, Jr. 2016). In other words, one of the ways that unjust hierarchy could be implemented in a legal system – especially in a common-law-derived system such as that of the United States – was to exclude the local knowledge and reasoning processes characteristic of subordinated groups from the adjudication of cases involving them.

This is a point that directly relates to the problem of platform colonialism discussed in Chapter 2. Legibility is a moral as well as a practical problem. When Twitter decides that an exception for enforcement action will be made for "legitimate public interest,"[27] the concept itself assumes a reference to an identified public whose interests are at stake that might be wholly distinct from the people in San Francisco who ultimately make the decision. In effect, in an enforcement content, "legitimate public interest" functions sort of like "reasonable person" in the law of negligence, that is, a reference to a community whose judgments the decisional process is meant to replicate. And in the context of global inequality, in which many people in other countries rightly perceive that Americans do not understand or respect their political processes and cultural values, the need to make judgments about things like "legitimate public interest" generates a demand to include the public whose interest is being legitimated (or not) in the decisions and their implementation. When we just let Americans make all the decisions, we end up with the sorry spectacle of American Facebook lawyers appealing in an American court to the superior wisdom of the US Congress in interpreting American free speech values in deciding whether or not it turns over evidence of a genocide in Myanmar to The Gambia to use in the International Court of Justice.

[26] See references at Kemp (1999, 967–68). For a particularly interesting discussion focusing on the negligence context and discussing other contrasting areas, see Gergen (1999).

[27] Cf. Twitter rules, enforcement philosophy, https://help.twitter.com/en/rules-and-policies/enforcement-philosophy, visited July 27, 2021; https://perma.cc/LP4V-U5ZD.

Moreover, work in fields like pragmatist epistemology can lead us to believe that colonial knowledge is also defective knowledge in terms of instrumental goals. For example, writing about moral philosophy – but with arguments equally applicable to corporate reasoning about things like content moderation – Elisabeth Anderson (2015) argues, drawing on classical pragmatists such as John Dewey, that reasoning processes under contexts of social hierarchy are tainted by the self-serving biases of the powerful, who have psychological dispositions to justify their own places in society. If this is right, and it certainly seems plausible, then it would seem that platform decision-making dominated by Americans and citizens of other wealthy liberal democracies would contain a built-in bias toward interpreting the needs of other societies in terms of the ideologies that they associate with their own success –for example, the alleged advantages of aggressive enforcement of intellectual property rights or less aggressive and facially neutral or "colorblind" enforcement of hate speech restrictions. This brings together the question of competence in terms of interpretation and the questions of morality in terms of inclusion: The same processes of inclusion that can make it possible for platforms to understand what a given utterance of maybe-hate speech means can also legitimate their judgment in determining what to do about it.

3.4 THE POLYCENTRIC MECHANICS OF DECENTRALIZED PARTICIPATION

The Myanmar case that leads this chapter focuses on enforcement – hate speech and incitement were already forbidden on Facebook, but lack of knowledge hampered enforcement efforts with tragic consequences. However, ordinary people on the peripheries also have useful knowledge about local needs that can bear on the making of the rules that are to be enforced. Human behavior generates endless novelty, including, alas, endless ways to cause harm to others or to shared resources; the effectiveness of policy solutions to shared problems can also depend on access to idiosyncratic knowledge as well as the epistemic benefits of a diverse community of reasoners (Cf. Hong and Page 2004). The high-speed evolution of behavior on the internet as well as the diversity of individual and social vulnerabilities suggests that the capacity of policy to adapt to novel behavior will require the ability to deploy knowledge far on the periphery from the people in San Francisco who currently write the rules for contemporary networked platforms.

Some scholars have, for that reason, advocated the aggressive decentralization of social media content moderation. For example, Wilson and Land (2021, 1074) argue that companies ought to create local teams in each country to write and enforce local rules.[28]

[28] To some extent, with the slow-boiling fall of Twitter, we may currently (as of this writing) be seeing experiments in far more radical decentralization, in which the federated Mastodon network creates (in effect) a market for moderation among interoperable microblogging platforms.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

I agree with Wilson and Land, but they perhaps do not go far enough. Company employees are still fundamentally implementing company norms and are still by some necessity socially distant from actual users and have at least partly conflicting interests. Moreover, discrete company employees are readily identifiable, and hence potentially vulnerable to government coercion, in particular in the countries where there are the most dangers of severe human rights violations of the sorts that Wilson and Land are concerned about, such as the Myanmar genocide.

One variation on the democratic strategy to mediate knowledge problems in the context of scale and diversity has, in the municipal and resource governance literature, gone by the name of "polycentricity." Associated with Elinor and Vincent Ostrom and the Bloomington school of political economy, polycentric organizations of government emphasize multiple levels operating at different scales with a degree of independence from one another as a method of generating policy innovation.[29] The notion of polycentricity seems to me to gel nicely with Ober's analysis of Athenian government, which emphasizes the capacity of democratic institutions that integrate diverse citizens with multiple sub-city affiliations and identities in multiple interactive institutional arrangements. In effect, the various kinds of possible governance actions of the Council, the Assembly, individual magistrates, and the citizen juries may have added this kind of polycentric character to the Athenian system. This, in turn, may have created a second level of incentives for knowledge aggregation, as citizens occupying these different institutional roles had to be brought into interaction with one another to reconcile their decisions, and citizens carried out different kinds of institutional roles calling for different kinds of knowledge and interactions when serving as a member of each of these groups.

Thus, it may behoove the designers of platform governance systems to consider what a polycentric system (or set of systems) of platform governance would look like. For example, could linguistic and subcultural communities on highly diverse platforms be given some regulatory autonomy, in order that they may both innovate in rules *and* transmit information to the center about developing idioms?

Several key insights from polycentrism theory are particularly relevant. The key proposition of the theory is that the grim predictions of the "tragedy of the commons" account of public resources, according to which users of a commons tend to defect from efforts to fairly allocate and limit overuse of that resource in the absence of external enforcement (or private property rights) (Hardin 1968), do not always empirically hold. Rather, in the real world, we see people successfully managing common resources all the time, if certain favorable conditions hold.

While platforms may or may not formally meet the definition of a common pool resource, they certainly share key features with them. In particular, they share the core incentive dilemma that users are often inclined to overuse and hence deplete

[29] See, generally, Aligica and Tarko (2012).

shared resources. For example, a user who posts "clickbait" on a social media site effectively overuses a common pool of attention and credibility that is shared by other users who also desire engagement with their content.[30] Hence, it may be useful to draw on this research in understanding the likely structural conditions of successful platform knowledge and incentive management.[31]

One of the key favorable conditions is the existence of mutual trust in shared compliance with joint resource-management arrangements (Ostrom 2009, 11). As Aristotle recognized first, however, this kind of trust is more difficult in larger scale governance organizations.[32] At the same time, large-scale governance organizations do a better job of managing externalities in resource governance, because they are more capable of including all people who might be affected by both positive and negative externalities within the ambit both of collective methods of decision-making and of enforcement authority.

This tension creates obvious problems for local resources that have significant externalities, or where individual actions have both substantial and important local externalities and still significant but less substantial distant externalities. Consider environmental regulation: The most serious consequences as well as benefits to pollution may be within a single city, as the residents of that city capture most of the economic benefits from local industrial activity, have to breathe most of the particulates spewed out by it, and so forth – however, activity in the city may also contribute to pollution problems over a much broader area.

As the name suggests, the solution proposed by theorists who focus on polycentricity – which has been particularly influential in environmental governance – is, roughly speaking, "both/and." That is, governance organizations on the local level

[30] Benkler, Faris, and Roberts (2018), 288 analogize – accurately, I think – clickbait publishers on social media to polluters; cf. Morell (2014) who interprets at least one early social media platform – the photo sharing service Flickr – as a knowledge commons in the sense given by Frischmann, Madison, and Strandburg (2014). Kiesler et al. (2011), 129–130 helpfully articulate a number of Ostrom-derived design principles which even today platform operators ignore at their peril. For example, they recommend rate limiting, Kiesler et al. (2011), 151, 136–38, which companies many years later rediscovered through, for example, limits on WhatsApp forwarding to reduce virality; Hern (2020).

[31] In the existing scholarly literature, Frey, Krafft, and Keegan (2019) argue, based on Ostrom's work, for the role of participatory design in regulating platforms. They bring together the notion of a "constitutional layer" for "digital institutions" (a concept including, but broader than, platforms) with a "participatory design" tradition in human–computer interaction to argue for the advantages of low-level (i.e., ordinary person, or ordinary user) participation in the basic governing processes of such institutions (including rulemaking). The argument for these design decisions rests on ideas similar to those discussed in this book thus far – the authors observe, for example, that Colorado's cannabis regulation software failed to adapt to its environment because of the state's inability to receive information and feedback from the peripheries. Another key recent work is Forte, Larco, and Bruckman (2009), who identify the de facto devolution of certain kinds of rulemaking (such as stylistic norms) from central Wikipedia into "Wikiprojects" as an example of Ostrom-esque polycentricity.

[32] Aristotle suggested that the maximum size for a city was constrained by the capacity for all citizens to know one another's characters (Aristot. Pol. 7.1326b). Ostrom (2010a, 661) similarly points out that trust is easier to achieve in face-to-face contexts.

as well as on higher levels are called for. Moreover, at the local level, theorists of polycentricity have identified that adjacent and even overlapping jurisdictions over people and activities may be beneficial rather than harmful, insofar as such arrangements provide for greater learning and competitive benefits: People under one jurisdiction can learn from the experiences of the next, and under some circumstances (such as in municipal service provision), jurisdictions can helpfully compete with one another to improve outcomes (Ostrom 2009, 33–34).[33]

Hiller and Shakelford (2018, 23–25) helpfully summarize the idea behind polycentric resource governance:

> Such a system is said to be polycentric if it has "many independent elements [that] are capable of mutual adjustment for ordering their relationships with one another within a general system of rules." Another definition of polycentric governance includes this concept's emphasis on "many decision centers having limited and autonomous prerogatives and operating under an overarching set of rules." A polycentric system, therefore, is not dependent on top-down government regulation, although it can include regulatory aspects. Instead, it is an organization of actors and method of governance that is multilayered yet interactive, independent yet networked, reinforcing and complementary. A polycentric system is complex, with differing rules and norms depending on the domain, providing for interaction between layers and among participants and allowing for "mutual monitoring, learning, and adaptation of better strategies over time." […] Professor Ostrom also emphasized organizational spontaneity, meaning "patterns of organization within a polycentric system will be self-generating or self-organizing," and "individuals acting at all levels will have the incentives to create or institute appropriate patterns of ordered relationships." Factors necessary for self-ordering include freedom to participate, rules, enforcement, and adaptation.[34]

A background paper on climate change which Elinor Ostrom (2009) wrote for a World Bank report is particularly instructive, as platform governance seems importantly similar to the way she describes climate change governance. In particular, in both contexts, unwanted conduct creates both local and global externalities, and hence can – at least in principle – be managed by both global and local actors.

Ostrom (2009, 16) notes that communities can act to mitigate air pollution in the aid of their own local air quality, and hence the incentives to address local externalities can also help mitigate global externalities.[35] Now consider that within local

[33] One digital example comes from the Forte et al. (2009, 63) paper, who note that when various wikiprojects have overlapping jurisdictions, it serves as an opportunity for discursive dispute resolution.

[34] Hiller and Shackelford (2018, 24–25), quoting, variously, Koontz et al. (2015); Aligica and Tarko (2012); Ostrom (2010b). The reader may notice that the story of interdependence and effective governance through bottom-up development in this summary sounds remarkably similar, at a high level, to Jacobs's understanding of a successful neighborhood.

[35] In the context of social media, of course, "local" is a bit of a shaky concept. Geographic localities are obviously relevant, but so are what, for lack of a better word, we might call "affinity localities" like #BlackTwitter – groups connected by a shared identity as well as a relative density of network

groups on social media, efforts to mitigate their own vulnerability to low-quality content might have beneficial spillover effects. For a concrete example, a number of the 2016 Russian attacks on US politics via social media were locally focused, such as an infamous protest in Houston instigated by the Internet Research Agency (Riedl et al. 2021). Such events cause local harms to the political groups thus manipulated as well as broader political harms; it may be possible for platforms to provide greater affordances permitting, for example, people invited to local events to verify the identities of other participants and the organizers in order to permit local policing of such inauthenticity.[36]

Similarly, Ostrom (2009, 23–27) notes that in the climate change context, large-scale efforts to control pollution have often had difficulties recognizing local variation. But a significant part of the lesson of this chapter is that large-scale efforts to control platform information pollution have also failed to handle local variation. Thus, we can borrow Ostrom's suggestion that it may be beneficial to promote both local regulation and global regulation, to obtain the benefits of interpersonal trust and communication as well as learning in local regulation and permit the knowledge generated by localities to propagate out to the global level through, for example, observations of what works on a local scale and local selection into effective regulatory arrangements.

However, the literature also suggests that such cross-scale interactions must be designed appropriately, they cannot be left to chance. For example, creating overarching governance arrangements to permit lower-level actors to interact also introduces the risk of suppressing the system's ability to accommodate local variation (Young 2002, 283). Currently, we might understand platform governance as characterized by harmful interactions between governance scales. Companies generally start with a Silicon Valley-libertarian approach to user behavior both on the social media side (in terms of "free speech") and on the transactional side (in terms of openness to all comers, for example, shoddy products purveyed by drop-shippers with dubious advertising claims); with ad hoc changes to rules as they discover behavior which either harms their interests or as they are subject to pressure from

interconnections or clustering which have for that reason similar properties to groups of people who share physical space offline. In this chapter, when I say "local" I mean both the geographic sense and the affinity sense unless there is some obvious reason to limit the discussion to one or the other.

[36] Thus, Riedl et al. (2021) report that nobody – participants, counterprotestors, or journalists – could figure out the identity of the leaders of the protest in Houston. While the Houston example is problematic because the IRA-organized event was a white supremacist protest in front of a Muslim center – and we presumably do not want to make it easier for white supremacists to organize and manage their internal affairs – the IRA also notoriously targeted much more sympathetic groups and ideologies, such as Black empowerment and liberation. See Senate Intelligence Committee Report on Russian Active Measures Campaigns and Interference in the 2012 U.S. Election, v. 2, www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf, 38–39. I give the Houston example particular attention simply because the Riedl et al. study gives us the most insight into the failures of local self-defense against this manipulation.

governments. Governments attempt to control conduct affecting their territories in a variety of ways, such as by controlling political speech which they do not like or defending intellectual property rights against counterfeiters on transactional platforms, but due to the frequent extraterritorial source of the behavior they wish to regulate, they typically find themselves operating via pressuring platforms rather than regulating directly. And users have little to no direct self-regulatory capacity, with the exception of some surfaces within social platforms with built-in self-regulation affordances (such as Discord groups and subreddits); likewise there is no international infrastructure to permit the conflicting claims of states to be effectively reconciled, or to provide for any formal engagement with civil society, indigenous peoples, and other organized nonstate actors outside of legislative processes.

At a minimum, experiments in what some commons scholars have called "co-management," in which higher-level and lower-level entities are jointly responsible for governing a resource (e.g., Berkes 2002), might involve, for example, some degree of devolution to groups of actual users of sufficient power over platform regulations to permit them to have some meaningful negotiating power with platform companies and governments. A part of the purpose of Chapter 6 is to suggest one way of implementing such an idea.

A closely related approach appears in a deeply insightful recent paper by political theorist Jennifer Forestal (2021a, on similar lines, see also 2021b, 2017). In it, she gives what amounts to a Deweyian version of the polycentrism theory, but focused less on governance and more on the structure of discourse on social media. Forestal argues, based on the theory of propaganda, that vulnerabilities to disinformation inhere in the underlying social structure of a population; in particular, its failure to be organized into salient but overlapping social groups which permit people to *both* engage in discussion on terms of shared interest and critically interact across those interests. While she does not use the term, the concept of "social capital" seems to me to serve as a good shorthand for such a propaganda-resistant social structure. Forestal compares Facebook's News Feed plus groups structure to Reddit's subreddit structure, ultimately arguing that the latter appears to be more successfully at promoting this sort of social capital.

In this context, one important lesson from bringing Forestal and Ostrom together is a recurrent point from the discussion in Chapter 1 of the way that the phenomenon of virality reveals a kind of false division between governance and product design. For the same overlapping but discrete groupings which might permit users to self-govern the content which they are producing and consuming can also, on Forestal's entirely plausible argument, improve their resilience against it in their capacities as mere users as well as democratic citizens. In view of Ostrom's insight, noted above, that polycentric governance tends to be "self-generating," we might hypothesize that a platform that is organized to facilitate the growth of social capital in terms of internal discursive health would also tend to promote the growth of groups suitable for a formal governance role. Hence, for example, existing subreddits might be recruited

to have more formal governance roles. An example of how such a development might come about is described in Chapter 6.

3.5 A DESIGN CRITERION FOR ANY POLYCENTRIC PLATFORM SYSTEM: ADAPTIVE CAPACITY

In addition to *distance* (in a geographic, cultural, or network sense), knowledge must also be managed over different *timescales*. Well-recognized strategies of governing in complex environments recognize that reliably predicting the outcomes of governance systems is often impossible in part because change is too swift, and that a core design criterion of any such system is to build the capacity for swift error correction in from the start.

An emerging area of social science and ecology that may characterize interactions on platforms – particularly, but not necessarily exclusively, the social media kind – is the notion of a complex adaptive system.[37] The "complexity" in such systems refers to the notion that assessing the entire system becomes difficult, including assessing the causal relationships between individual elements (or interventions on them) and the higher-level properties or other elements; actions can have "nonlinear" or "emergent" consequences (Miller and Page 2007, 3, 10, 27–28).[38] Miller and Page (2007, 233) observe that complexity tends to appear in systems that feature "heterogeneity, adaptation, local interactions, feedback, and externalities," which seem to me to generally characterize behavior on social media networks, and perhaps platforms as a whole. On such platforms, there are numerous actors, both institutional and individual (including ordinary users, companies, and even nation-states), who shape their behavior partly in response to the responses others give to their behavior, and so forth, in multi-directional feedback systems. One example is the perennial arms race between the "search engine optimization" industry and Google, as well as the ways in which content producers and consumers respond to changes in social media feed algorithms. Such feedback effects only increase when platforms interact with other social systems like markets or elections, as when, for example, r/WallStreetBets created massive shifts in the price of GameStop and other stocks, until controversial industry responses put a stop to it (Mezrich 2021). Heterogeneity and externalities are obviously present in such systems for the reasons discussed in the rest of this book.

This suggests that the choice of institutions to facilitate such governance must be particularly attentive to challenges involving understanding both local and global

[37] Cf. Lymperopoulos and Ioannou (2016), arguing as much.

[38] I understand "nonlinear" in this context to refer to a kind of disproportion rooted in the conjunction of feedback loops and extensive interconnections among agents, in which one seemingly small intervention can lead to chains of adaptive responses from the underlying agents, and hence surprising consequences, including in distant locations. "Emergent" describes "aggregate properties that are not directly tied to agent details" (Miller and Page 2007, 53).

effects of interventions (including seemingly small interventions), managing those effects across diverse sets of actors, and adapting to the discovery of surprising behavior – both positive and negative – in response to those interventions.

Within a polycentric governance framework, another potentially useful theoretical construct drawn from ecosystem scholarship that particularly focuses on complex systems is the notion of adaptive management. In a recent article on the management of the boundary between urban and wilderness areas, Craig and Ruhl (2020) helpfully describe both the contexts in which such strategies are likely to be most useful and their basic features. On their account, adaptive management is a "formal, structured decision process [which] involves a 'setup' phase, during which the decision-making actor specifies stakeholder involvement, management objectives, management actions, models, and monitoring plans, followed by an 'iterative' phase, during which the actor specifies the decision-making process, follow-up monitoring, assessment, and feedback" (Craig and Ruhl 2020, 616). That system is "resource-intensive" in view of the continuing infrastructure for monitoring and decision-making that it requires, especially in light of the fact that "external stakeholder engagement" is necessary to aggregate the kinds of information necessary for adaptation, but at the same time must be carefully crafted to avoid the risk of "suffocat[ing] the iterative decision process" by imposing excessive impediments to change (Craig and Ruhl 2020, 616–17).

Adaptive management is appropriate for systems that have several properties, including "chang[ing] dynamically over time," high uncertainty about the harms which may be caused by interventions on them, high "controllability," that is, there are lots of manipulations we can actually engage in, and a low risk of "irreversible transformation" (Craig and Ruhl 2020, 616). Under such circumstances, learning-oriented management strategies are called for in view of the fact that such strategies permit quick iteration (accommodating the changing nature of a system) in a context where such iteration is possible (because of high controllability) and minimally harmful (because changes are not irreversible); such iteration ameliorates the characteristic problem of such systems, namely the difficulty of predicting consequences, essentially by experimenting and seeing what happens. One characteristic type of system which is likely to meet the criteria described above is a complex adaptive system, one in which interdependent elements with feedback effects among themselves lead to emergent patterns of behavior and structure (Craig and Ruhl 2020, 614).

Because platforms probably constitute complex adaptive systems, and because a lack of information as to what is going on at the fringes of a platform ecosystem – and hence, by implication, as to the consequences of platform design choices and interventions on governance processes and rules – is a core challenge in platform governance, adaptive management may offer a particularly plausible toolkit to thinking about platform governance.

One challenge for the notion of adaptive management is that catastrophic consequences to experimental changes may indeed be possible – certainly existing design features of social media platforms have led to or contributed to catastrophic

consequences, such as the genocide in Myanmar and electoral distortions in the United States, Britain, and the Philippines. However, this may not entail the abandonment of adaptive ideas, but rather suggest that any form of learning-oriented governance must come with guardrails, such as a bias toward discovery of the sorts of information likely to point to catastrophic consequences in time to intervene ahead of those consequences. Moreover, given that our existing social media platforms are *already* leading to catastrophic consequences, if it is true that predicting the effects of interventions is impossible (because of complexity), then it may be that the probability of catastrophe would be reduced under a higher rate of experimental interventions (or it may not be; it cannot be known).[39]

Another consideration at play in deliberating on the adoption of adaptive management as an explicit platform governance strategy is the role of learning capacity within the underlying system. As Doremus (2011, 1471–72) argues, not all enterprises of governance offer a meaningful prospect for learning; in particular, if feedback on governance experiments is insufficiently fast, learning may occur insufficiently quickly to permit knowledge to be acted on. In the platform context, a related worry may arise, namely that due to the scale and diversity characteristic of platforms, the environment in which governance decisions are undertaken may change so quickly that, even if learning is possible, learning isn't fast enough to permit adaptation before it is, in effect, mooted by new environmental changes (cf. Doremus 2011, 1474 on the possibility of confounding changes in the environment). This suggests that any adaptive governance interventions must be accompanied by close attention to the speed at which knowledge is moved to the site at which adaptations are required. In the system described in Chapter 6 of this volume, this may suggest a bias toward giving local governance groups or councils more direct control over local platform outcomes, in order to permit swifter incorporation of new information close to the site of its impact.

In complex systems, adaptive governance arrangements may emerge rather than being imposed in the top-down fashion described by Craig and Ruhl; one way to represent the task of a governor in such a system is to create the conditions under which such adaptive institutions may emerge in ways that facilitate the achievement of public goals (Cosens et al. 2021). As Cosens et al. recognize, there is an overlap between the ideas of adaptive governance, new governance (which, as described in the introduction to this volume, focuses on multistakeholder inclusion), and Bloomington School polycentric solutions.[40] In their words:

[39] One plausible way to model the problem is as a landscape of possible platform design and governance configurations, some proportion of them which pose catastrophic risk; it may be that we are currently in a region of that landscape in which many configurations lead to such risks, but that other regions lead to lower risks. But this depends on one's priors on the distribution of catastrophic risk, and there does not seem to me to be much to recommend any particular view.

[40] I read the shift in emphasis over some of these papers from "adaptive management" to "adaptive governance" as perhaps partly reflecting a greater bottom-up character – however, I might just be over-reading a minor linguistic difference.

> New governance and adaptive governance both include as essential features bottom-up self-organization and collaboration; public, private, and public–private networks; and multiple nested centers of authority (i.e., polycentricity). Adaptive governance literature calls out the presence of processes to manage uncertainty and mechanisms for learning and incremental adaptation. Both emergent forms of governance respond to increased connectivity and complexity and the corresponding need for contextualization and adaptation. They appear able to navigate and maintain overall social stability when system trajectory is uncertain and fraught with surprise. In short, they represent societal responses to complexity.[41]

In many ways, this version of adaptive governance which offers specific attention to the notion of preserving the public character of multistakeholder models while consciously adapting to social complexity, sounds remarkably like a reprise of John Dewey.[42] Elizabeth Anderson (2006, 13–14) describes Dewey's conception of democracy as analogous to the scientific method, in which public policies are framed in deliberation among diverse and inclusive groups and then tested to determine their outcomes, with democratic institutions like free speech and periodic elections providing information both for hypothesis generation and feedback/observation. What the idea of adaptive governance adds is a healthy skepticism of predictive methods of policy framing, and hence a bias toward governance systems that can generate and change ideas and experiments at a higher rate and are more responsive to feedback. In a complex system with actors across multiple locations, the vigorous use of democratic polycentricity may, I hypothesize, facilitate learning in the same way that federalism understood as the creation of political laboratories does, that is, by creating additional sources of change and sites of observation (with measures to swiftly disperse those observations across the governing network) while permitting simultaneous experimentation with novel options.

Chapter 6 of this book will sketch out a system of bottom-up platform governance designed with the capacity to adapt to change as a first-order goal. But before we get there, we must first look in the opposite direction: Governance requires not only innovation but also constraint, and many of the problems generated by platforms are arguably attributable to the lack of constraint, particularly of company executives. Chapter 4 considers the ways in which political states facilitate that constraint, so that we may later design institutions that can meet both needs: Innovation from below and constraint at the top.

[41] Cosens et al. (2021, 4), internal citations omitted.
[42] Doremus (2011, 1464 n. 35) notes that a number of prior scholars have read adaptive management in the Deweyian tradition.

4

The Problem of Platform Self-control

The platform companies knew, or should have known, that their platforms were being abused during the Trump administration. Unlike in Myanmar, the constant sharing of misinformation and incitement of hate associated with Donald Trump was going on in English, and in the United States – the language and the country best covered by the content moderation tools of every company. It is well established that Facebook in particular realized at the latest shortly after the 2016 election that a small army of trolls, partially in the pay of Russia, was working in favor of Donald Trump (Madrigal 2017).

In response to this knowledge, according to whistleblower Frances Haugen, Facebook undertook measures during 2020 to protect the public from the spread of particularly dangerous information, such as false claims that the election had been stolen (Looft and Ferris 2021). Unfortunately, as Haugen reports, those measures were removed prior to January 6 and were not reinstated until after the coup attempt (Looft and Ferris 2021).

To some extent, the failure to control the spread of election-related lies simply resulted from challenges with enforcing company policies at scale. For example, according to a leaked internal Meta document, events moved too swiftly for company investigative personnel to identify that the various "stop the steal" and related groups were working in concert (Mac, Silverman, and Lytvynenko 2021). Since Facebook is the largest social media company and accordingly had the most resources to do this work, it's hard to imagine that any other company could do better.

And yet, technical and investigatory challenge is not the full story, for there's also plenty of evidence that the company was scared away from addressing right-wing misinformation. During the Trump administration, for example, it was clear that far-right "news" publisher Breitbart was a major source of misinformation. According to another whistleblower, when company employees asked why nothing was being done, Joel Kaplan, the head of Facebook's policy team, replied: "Do you want to start a fight with Steve Bannon?" (Timberg 2021).

I claim that this fear of retaliation by powerful political figures, combined with the temptation to short-term profit and some degree of internal political conflict,

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

drives a problem of self-control that can be seen across the platform context. This chapter aims to offer some solutions for it.

4.1 MARK ZUCKERBERG ISN'T PLOTTING TO FIX THE ELECTION FOR YOUR POLITICAL ENEMIES, I PROMISE

On May 16, 2020, the then-President of the United States alleged – ironically, on Twitter itself – that "The Radical Left is in total command & control of Facebook, Instagram, Twitter and Google. The Administration is working to remedy this illegal situation. Stay tuned, and send names & events."[1] A year beforehand, Trump had alleged that Twitter "make[s] it very hard for people" to follow him, and that "[i]f I announced tomorrow that I'm going to become a nice liberal Democrat, I would pick up five times more followers."[2] According to anonymous sources, Department of Justice officials in the Trump administration were toying with proposals to hold social media companies liable for alleged political censorship before he left office.[3]

Around the same period, Senator Josh Hawley (R-MO) proposed an "Ending Support for Internet Censorship Act," which, in the words of the press release announcing the bill, "removes the immunity big tech companies receive under Section 230 unless they submit to an external audit that proves by clear and convincing evidence that their algorithms and content-removal practices are politically neutral"; the bill was allegedly in response to "a growing list of evidence that shows big tech companies making editorial decisions to censor viewpoints they disagree with."[4] A month after Hawley, Representative Paul Gosar (R-AZ) proposed a similar bill.[5]

These allegations and proposals should be puzzling to a student of American capitalism. After all, social media companies are under a fiduciary obligation to try to produce profits for their shareholders. Organizing themselves on the basis of ideology would alienate – expensively – many of their users, plus the government in any Republican administration in the United States and similar regimes in many other countries. There is a reason that we don't see many US publicly traded companies going all-in on a particular political party. Unsurprisingly, Facebook, Twitter,

[1] Donald J. Trump Twitter May 16, 2020, https://twitter.com/realDonaldTrump/status/1261626674686447621.

[2] Margaret Harding McGill and Cristiano Lima, "White House to Hold Social Media Summit amid Trump Attacks," Politico, June 26, 2019, www.politico.com/story/2019/06/26/white-house-social-media-summit-1383280.

[3] Tony Romm, "Attorney General Barr Blasts Big Tech, Raising Prospect That Firms Could Be Held Liable for Dangerous, Viral Content Online," *Washington Post*, February 19, 2020, www.washingtonpost.com/technology/2020/02/19/attorney-general-barr-blasts-big-tech-questioning-its-protection-liability-content/.

[4] "Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies," June 19, 2019, www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies; S. 1914.

[5] H.R. 4027; "Congressman Gosar Introduces Legislation to Stop Big Tech Censorship," July 25, 2019, https://gosar.house.gov/news/documentsingle.aspx?DocumentID=3854. For a description of numerous other efforts to amend or repeal Section 230, see Samuelson (2021).

and YouTube deny that there is any political censorship going on their platforms (with the exception of some Musk claims about prior Twitter practice), and no partisan criteria can be discerned anywhere in their published content rules. Thus, conservatives in the United States who accuse them of censorship typically do not accuse them of having explicit and formal partisan policies – mainstream conservatives do not claim that there's a "no Republicans" rule somewhere in Facebook's Community Standards. Rather, the basic theory seems to be something like that company employees – located, after all, primarily in notoriously liberal Northern California – are bringing their own political ideologies to their policy enforcement decisions in the context of ambiguous rules and nontransparent enforcement systems (e.g., McGregor and Kreiss 2019).[6]

But this isn't just some kind of right-wing conspiracy theory. The left makes similar allegations, and they are just as puzzling, if slightly different. The left, rather than accusing the companies of over-enforcing their content-moderation rules in a politically biased way, accuses them of under-enforcement against right-wing violations of those rules. For example, then-Senator Kamala Harris sent a letter to the Twitter CEO recounting several of Trump's tweets, alleging that they violate the Twitter terms of service, and seeking the removal of Trump's account more than a year before his account got removed for the January 6 autogolpe bid.[7] Persistently, commentators on the left have accused the companies of running scared from threats of regulation by the right – or simply of being more interested in the profits to be gained from the engagement that people like Donald Trump generate than in the safety of their other users and the country – and hence failing to control harassment, hate speech, and dishonest propaganda that violates their own policies.[8]

The left-wing critique makes marginally more economic sense, at least in the short-term, than the right-wing one. Conceivably, a racist or violent political leader could drive enough profitable engagement that it might be in the interest

[6] Perhaps there is also a secondary claim that some platform rules themselves, such as prohibitions on hate speech, are inherently biased. This claim isn't made so frequently or loudly, I guess because it's still considered impolite to admit that one's political ideology entails racial slurs and the like.

[7] Letter from Kamala Harris to Jack Dorsey, October 1, 2019, https://assets.bwbx.io/documents/users/iqjWHBFdfxIU/r3ukxGLqQFLA/v0 (last visited March 14, 2020) (see also Ziady 2019). Twitter has carved out a formal exception for people like Trump. Twitter, "About Public-Interest Exceptions on Twitter," https://help.twitter.com/en/rules-and-policies/public-interest [https://perma.cc/GA5Q-FG2C] (last visited December 4, 2022), describes an exception to the Twitter rules for policy violating content by elected officials, and says that such content will be put "behind" a "notice" of its rule violating character.

[8] Even after Trump was banned, some commentators suggested that Facebook's particular approach to the ban was motivated by advertising interests. For example, Ryan Goodman of NYU, commenting on the Oversight Board's upholding of Trump's Ban: "Did Facebook have financial incentive to avoid permanently banning Trump? This in @OversightBoard decision struck me: Facebook refused to answer Board's question 'whether account suspension or deletion impacts the ability of advertisers to target the accounts of followers'." post of May 5, 2021; https://twitter.com/rgoodlaw/status/1389954225409601538.

of Facebook's or Twitter's shareholders to leave their hate speech alone. But that is true only if the leaders of those companies heavily discount the future. After all, it's no coincidence that Google's content regulation team is called "trust and safety": A social media platform filled with hate speech, threats of violence, or harassment is likely in the long term to drive away its users in fear or horror, either to competitors or to no platform at all.[9] (For reasons noted later in this chapter, this user flight is likely to come by surprise and all of a sudden.) Platform leaders know that, so the left-wing critique seems to assume that they're willing to sacrifice the long-term security of their user-base for a few extra clicks in the immediate present. Yet these same leaders have a well-known predilection to make very large, very long-term investments, such as Alphabet's investment in its famous "moonshot" division, X, and Facebook's creation of "Internet.org" (thereafter "Free Basics") to expand internet connectivity in developing markets, plus its huge acquisitions made at a staggering cost in order to increase the number of users it can reach without an immediate revenue source, most famously its acquisition of WhatsApp. This is not the behavior of corporate executives with a short-sighted focus on the next quarter's profit and loss statements.

The potential flight of their users is not the only way that toxic content harms platform interests. As Gillespie (2018a, 14) recounts, there has at least been (plausible) speculation that the toxicity of Twitter's environment scared off possible corporate acquirers years before Elon Musk bought it. Moreover, many advertisers, concerned with "brand safety" – that is, with avoiding the risk of associating their brands with upsetting or disreputable content – are likely to flee platforms with too high a proportion of toxic content, even if adherents to one of the country's major political persuasions find that content less objectionable (Bellman et al. 2018; Kulp 2019; Braun and Eklund 2019, 6; for a chronology of YouTube's brand safety challenges, see Pottinger 2018–2019, 525–31).

In short, neither the censorship of innocent speech nor the refusal to remove harassment and propaganda is likely to be consistent with the long-run economic interests of social media companies. If those companies can actually control their own behavior, both the left-wing and the right-wing complaints fail the test of plausibility. Yet, as I will discuss further in a moment, there is a little bit of evidence that some of the complained-of behavior is actually happening. How? I claim that, to the extent either complaint has any grounding in fact, what really has to be going on is a kind of failure of *corporate self-control*, a succumbing to short-term temptation to squeeze out a few extra clicks or satisfy a few noisy left-wing engineers or vengeful politicians at the expense of long-run company interests.

There is some evidence supporting the complaints of both sides in the US political context. For example, one detail prominent in many of the right-wing complaints

[9] For an interesting recent economic analysis of the incentives underneath social media content moderation, see Liu, Yildirim, and Zhang (2021).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

is a practice known as "shadowbanning," in which a user isn't actually banned from the platform, but the content they produce is silently made invisible, or less visible, to other users. And while the companies deny political shadowbanning, it is uncontested that they have, and exercise, the power to shadowban – they admit to using reductions in the distribution of content as a lever for enforcing their policies.[10] Moreover, both employees of technology companies and wealthy technology entrepreneurs really do predominantly lean to the left, or, at least, support Democrats over Republicans (Broockman, Ferenstein, and Malhotra 2019). There was also at least one relatively concrete and high-profile case that could fairly be ascribed in part to political bias in content moderation on Twitter, where the company refused to allow the distribution of a New York Post story about Hunter Biden in what appears (though much is still unclear) to be an erroneous application of company rules about content derived from hacking (Vlamis 2022; Schreckinger 2022).[11]

Similarly, the left-wing complaint also benefits from some factual support: There are some credible press reports suggesting that Facebook officials in fact consciously interpreted their platform rules charitably toward conservatives in order to avoid an appearance of left-wing bias or offending conservative lawmakers.[12] More generally, there is evidence that at least some platform company executives were in fact motivated by short-term growth metrics as opposed to longer-term platform integrity. A whistleblower report to the SEC, FTC, and DOJ by Peiter Zatko ("Mudge") alleges that Twitter executives turned off measures meant to prevent "spam bots" because they were rewarded for increasing measures of active users.[13] The economics of platforms can undermine rule enforcement in other ways. For example, in some

[10] See, e.g., Facebook Community Standards #20: False News, www.facebook.com/communitystandards/false_news (last visited March 15, 2020), which states that "we don't remove false news from Facebook, but instead significantly reduce its distribution by showing it lower in the News Feed"; Twitter, "Our range of enforcement options," https://help.twitter.com/en/rules-and-policies/enforcement-options (last visited March 15, 2020), which include "[l]imiting Tweet visibility." Gillespie (2022) further describes shadowbanning and related phenomena. It's worth nothing that shadowbanning is the obviously correct solution in some cases. For example, the dating site Bumble reportedly uses a form of shadowbanning to attempt to get rid of users accused of sexual assault – from the user's perspective, their profile is still on the site, but nobody else can see it, hopefully making it harder for the "silent block[ed]" user to know they're banned and create another account or retaliate against the person who reported them (Edwards et al. 2021). Given the extreme harms inflicted by sexual assault, this seems like an entirely reasonable policy to me (even done on the basis of mere accusation, since an alleged assaulter's interest in being on a dating app isn't significant enough to be entitled to any substantial process before acting proactively to protect those with whom they might interact).

[11] Well-known technology industry blogger Mike Masnick has an overall summary of the events surrounding the Hunter Biden laptop leak at "Hello! You've Been Referred Here Because You're Wrong about Twitter and Hunter Biden's Laptop" (TechDirt, December 7, 2022), www.techdirt.com/2022/12/07/hello-youve-been-referred-here-because-youre-wrong-about-twitter-and-hunter-bidens-laptop/.

[12] Journalist Steven Levy (2020a, 340–45), who was given an unusual amount of access to Facebook personnel, reports on a number of instances in 2015–6 where Facebook was particularly solicitous to conservative fears of "censorship" even of rule-violating content.

[13] Whistleblower report of Peiter Zatko, https://s3.documentcloud.org/documents/22186782/whistleblower_disclosure.pdf, pp. 10–12.

companies the "adjudication" of rule violation is merged with a customer service function, creating inconsistent rule enforcement across revenue classes. Persistent public controversy about Meta's "crosscheck" system, which allegedly provided for special scrutiny before platform sanctions could be levied against prominent (and hence revenue generating) accounts is one example (Horwitz 2021). According to the Oversight Board's evaluation of that system, not only figures of public importance such as elected officials are included – so are "business partners" who are relevant to Meta's bottom line.[14] Another example is YouTube's tiered system of customer service for content creators, which Caplan and Gillespie (2020, 7) allege provides greater affordances to appeal platform sanctions for higher revenue users.

Again, I think the evidence is somewhat stronger for the left-wing complaint (though this evaluation may be colored by my own left-wing politics). It is consistent with evidence of similar company behavior outside of the country: There are credible allegations that Meta also soft-pedaled rule enforcement against the Modi regime in India (Purnell and Horwitz 2020). Allegations about Meta's failures in India are strikingly parallel to those in the United States: In both countries, a high-level company "policy" official – where "policy" is an organizational function simultaneously responsible for lobbying and government relations and for rule development and enforcement – allegedly intervened in rule enforcement in order to protect a government in power, with allegedly mixed motives to both promote that official's personal ideology aligned with the government and to shield the company from government retaliation.[15] One person came up with an experiment to tease out the inconsistency in Facebook policies: A Facebook user created a page entitled "Will they suspend me?," which quoted Trump's posts to see whether the same standard would be applied to an ordinary person (O'Kane 2020). The same standard was not applied.

Even fairly small platforms face serious problems with consistent rule enforcement. In 2021 OnlyFans, the amateur subscription video platform that became a

[14] Oversight Board policy advisory opinion on cross-check program, December 2022, https://oversightboard.com/attachment/440576264909311/, p. 8. See also ibid., p. 30, for an example, in which a celebrity appears to have been given favorable content moderation treatment through the cross-check system in anticipation of signing an exclusive deal with a Facebook streaming service.

[15] Compare the press accounts of the behavior of Ankhi Das (Purnell and Horwitz 2020) and Joel Kaplan (Mac and Silverman 2021). Another press report suggests that there may have been similar dynamics in Brazil and in the U.K. In Brazil, a "subject-matter expert" refused to permit the removal of a speech by Jair Bolsonaro in which he described indigenous Brazilians as "evolving and becoming, more and more, a human being like us" (Marantz 2020). But it turned out that the "expert" in question had previously worked for a political ally of Bolsonaro's. The employee who spoke to the press suggested that the refusal to remove the speech was likely also motivated by advertising revenue. That same report described far-right pages in the U.K. that were "shielded" – that is, excluded from ordinary rules requiring a ban of a page after enough content policy violations. According to the article, content moderators perceived that shielded pages "tended to be those with sizable follower counts, or with significant cultural or political clout – pages whose removal might interrupt a meaningful flow of revenue" (Marantz 2020).

center of the online sex worker trade during the COVID-19 pandemic, announced that it would ban pornography (as far as I can tell, the only thing that anyone has used it for). Unsurprisingly, this ban didn't last – less than a week later someone realized that doing so would destroy the company and found some way to make keeping the pornography work (Spangler 2021). But while the stated reason for the short-lived porn ban on the porn site involved pressure from payment processors and other financial intermediaries, one might reasonably suspect that some of that pressure to stop hosting pornography was related to evidence turned up in a BBC investigation that it had failed to enforce its own rules. It turns out that OnlyFans had been hosting unlawful conduct (in relevant jurisdictions) such as prostitution and, most alarmingly, sexual videos featuring people under 18 (Titheradge 2021). Making matters worse, the BBC reported (Titheradge 2021) that more popular (and hence profitable) accounts were given – as a matter of written policy! – more "warnings" before being shut down for violations of platform rules, including rules about illegal content. It sure seems like a failure to comply with their own rules or underlying law almost killed the company.

Nonetheless, while supported by some evidence, both the left-wing and the right-wing complaints about the major social media companies are almost certainly exaggerated. Some complaints may simply be due to differences in interpretation of platform rules; so long as companies are enforcing their rules in good faith, such differences in opinion ought not to be seen as unfair political bias. For example, Kamala Harris's letter to Twitter accuses Trump of violating Twitter rules prohibiting "harassment" and "the glorification of violence," and cites as examples tweets in which Trump intimated that a whistleblower in the executive branch was guilty of espionage. Arguably, those tweets, and the others that Harris cited (including a particularly menacing one about "a Civil War like fracture in this Nation" if Trump was successfully removed from office by impeachment) constituted harassment or threats of violence. In Harris's words, "These tweets should also be placed in the proper context, where the President has compared the whistleblower to a 'spy' who may have committed treason, and further implied that the punishment for that should be death."

However, an equally reasonable person could believe that the tweets were no such thing. "Harassment" is a notoriously slippery concept. And Trump's tweets, regardless of what one might think about their overall democratic propriety, could just as easily be interpreted as nothing more than the attempts of a politician facing serious accusations to defend himself by impugning the behavior of his accusers and by warning of the political (not violent) consequences of their actions. Similarly, the "Will they suspend me" disparity between enforcement against Trump and against an ordinary person can potentially be attributed to the existence of the infamous cross-check system – according to which prominent users weren't subjected to different rules but were afforded different *process* in the form of a second level of company scrutiny before their Facebook posts were taken down. According to the person behind "Will they suspend me," Facebook ultimately claimed that one of his posts was removed in

error (O'Kane 2020), and this is consistent with the notion that Facebook was less willing to accept erroneous removals of the content of prominent people like Trump.[16]

The examples of Harris's letter and "Will they suspend me" suggest what is intuitively obvious to most of us as I write these words at the end of 2022: The social media platforms are faced with pervasive distrust and skepticism relating to the fact that their rules are not enforced transparently and are also subject to substantial debate in their application. But faithful enforcement in accordance with platforms' own rules may be just as damaging to their bottom lines as inconsistency if the external world cannot observe that enforcement actually is consistent. If platforms cannot reassure users and regulators that they are neutrally enforcing their rules, then they might suffer a lack of user trust or retaliatory regulation regardless of the actual state of affairs inside corporate offices. To illustrate this phenomenon, we can note that, regardless of whether pre-Musk Twitter was actually censoring conservatives (e.g., whether that Hunter Biden incident was a good-faith mistake or colored by the partisan affiliation of the company employees who made the call), enough people on the right believed that censorship was happening that many fled to various extreme right wing "free speech" social media platforms such as Parler, Gab, and "Truth Social" (the last of which was founded by Donald Trump). "Gab," for example, competes with other platforms on the basis of its lack of "political censorship" (Zannettou et al. 2018; Lima et al. 2018). Twitter's loss of a sizeable chunk of the users from one political affiliation is a problem from the standpoint of a company whose revenues are tied to scale; it may also be a problem from the standpoint of society to the extent the flight of the far right into unmoderated echo chambers promotes their further polarization and radicalization. Thus, platforms need not only enforce their rules neutrally but also must convince their users to trust that they are doing so. Even if they can control themselves, in other words, they need to be able to control themselves *in public*.

4.1.1 *Sometimes Failures of Self-control Are Just Failures of Corporate Governance*

In some cases, an organization's failures of self-control do not reach its top ranks. Low-level employees might frustrate the policy choices of top-level leaders by disobeying the rules, or by distorting their application. For example, the frontline

[16] Such a policy isn't necessarily irrational or bad – it's problematic if purely motivated by revenue, but it might make sense for public discourse reasons – because of the greater attention paid to posts by particularly prominent people, it might make more sense to be more careful about disruptively removing and then restoring those posts. On the other hand, this argument only holds if the cross-check process happens quickly, otherwise prominent people could – as Trump in effect did – take advantage of the extra time it affords to use their gigantic audiences to do immense public harm. This delay is one of the risks identified by the Meta Oversight Board in its policy recommendations surrounding cross-check, see Oversight Board, "Oversight Board publishes policy advisory opinion on Meta's cross-check program," www.oversightboard.com/news/501654971916288-oversight-board-publishes-policy-advisory-opinion-on-meta-s-cross-check-program/ (December 2022).

workers charged with interpreting social media content rules might be interpreting – intentionally or inadvertently – ambiguous rules concerning "harassment" in such a way that conservative speech is seen as harassing at a different rate than liberal speech, independent of the inherent nature of such speech (if such a thing is conceptually coherent).

This is unlikely to be the case with respect to the most politically salient controversies, such as the debates over social media bans of particular prominent conservatives. Low-level employees are not quietly making the decisions about whether or not to ban Trump from Twitter or Alex Jones from Facebook.[17] However, to the extent general patterns of enforcement or nonenforcement are nonetheless skewed in less prominent cases, this could, in the aggregate, undermine both the capacity of these platforms to preserve the environment their leaders are trying to create and public trust in their neutrality.

Another example of the failure of corporate self-control through employee defection comes, not from social media political censorship and hate speech, but from Amazon's marketplace. Amazon owns a number of private label brands, such as "Amazon Essentials," which compete with the products of third-party sellers on its platform. Amazon also collects immense amounts of competitively valuable sales data on behalf of its third-party sellers. This creates an obvious conflict of interest: Amazon has a short-term incentive to effectively engage in industrial espionage against its own sellers. In the absence of some mechanism for corporate self-control, this conflict of interest could deter sellers from using Amazon to distribute their products; accordingly, the company has a policy of not using individual seller data from its third-party sellers in choosing which private label products to release. Recently, it came out that Amazon's employees violated this policy, exploiting loopholes in the firewall between the sales data side of the business and the private label side to do so.[18]

Amazon's failure to control employees resembles more general failures of employee management that have allowed companies to stumble into unethical conduct. Familiar cases of such employee malfeasance in the nonplatform economy include foreign sales employees succumbing to the temptation to bribe government officials and mortgage brokers writing "stated income" (a.k.a. "liar") loans. Hence, to some degree, such problems might be amenable to conventional management techniques such as tightening the enforcement of internal rules and conducting random audits.

However, because such internal controls are always imperfect, and Amazon is a platform, it also raises many of the same user trust issues as political "censorship"

[17] Similarly, emails that Musk released from Twitter relating to the Hunter Biden story noted above show the involvement of various senior Twitter personnel at the time, including its head of trust and safety and its general counsel.

[18] The details of this story come from Dana Mattioli (2020). Sam Bowman (2020) of the Adam Smith Institute has a helpful discussion of the incentive Amazon has to avoid misusing third-party seller data in order to avoid deterring product innovation on its platform.

on social media. Amazon does not only need to control its employees, it needs to make control of its employees credible to third-party sellers. This contrasts with other contexts of employee misconduct – for example, mortgage brokerages do not need to credibly signal to their customers that they have adequate internal controls to stop writing liar loans.[19] This illustrates that when employee and leadership goals diverge, traditional techniques of management do not exhaust the strategies of corporate self-control that are relevant; when there are external constituencies who need credible signals of trustworthiness to participate in a platform ecosystem, companies also require tools to make themselves externally accountable.

4.2 CAN PLATFORMS COMMIT THEMSELVES TO GOVERN CONSISTENTLY?

Lack of self-control doesn't come out of nowhere. We have fairly ordinary ways of thinking about its sources. For example, an organization may lack internal management capacity: Corporate leaders might issue directives but lack the power to make the subordinates who actually have their fingers on the metaphorical button that deletes a particular piece of content or bans a particular user comply with them. Or an organization's leadership might suffer from a lack of willpower: They might really want to act in a long-term profit-maximizing way but succumb to the temptation to leave up Alex Jones's conspiracy theory or delete some Republican's speech in the face of short-term profit or political pressures.

The problem of top-level leadership is worth further examination. Right now, the dominant framing of the problem of excessive power in social media content moderation is one of excessive company power. On this framing, companies are understood as monolithic entities with a unified will – understood, for example, from the American political right, to mostly be represented by the general political ideology of Northern California, hence the endless complaints that left-wing technology workers are censoring conservatives.

[19] However, the government may require them to communicate credible information about their internal controls so that it can economize on investigative costs. Also, to some extent mortgage brokers may need to credibly signal to lenders/purchasers of mortgages that they do not write liar loans, but given that there are many fewer mortgage lenders and buyers than there are Amazon sellers, and lenders and buyers are likely to have their own investigative resources and benefit from existing infrastructure such as third-party auditing (which is obviously imperfect, as the financial crisis taught us), the problems are less difficult than they are for a many-to-many platform like Amazon. This dynamic arguably also exists in other non-platform contexts. For example, airlines need to make sure customers trust their employees' compliance with safety regulations – although they have massive government regulation helping them to do so. In some platform contexts, companies might also welcome government regulation in order to have a third-party guarantor of their conduct. Amazon's interest, for example, would probably be served by such a regulation (combined with real auditing and enforcement) insofar as it could then tell sellers "we won't steal your data, because if we do, the government will impose massive fines on us."

Yet press reports as well as the accounts of former company employees, for example from the famous "Facebook Files" leak, suggest, to the contrary, that there are routine differences between company employees responsible for rule implementation and senior managers (Birnbaum 2021).[20] And the shape of those differences resembles a familiar problem in political theory: High-level executives are tempted to command deviations from general rules, either to meet short-term crises or to achieve short-term gains against the long-term benefit of rule enforcement.

To illustrate this dynamic, consider the saga of Donald Trump's Twitter and Facebook accounts. During Donald Trump's presidency, there was a widespread public debate about the extent to which his social media posts – like those of other far-right figures – violated platform rules. Part of the problem is that there was a certain degree of ambiguity with respect to those rules in the first place; in particular, at least until April 2018, significant parts of Facebook's content policy enforcement guidelines were not available to the public (Bickert 2018). However, there were certainly plausible arguments that Trump's behavior violated the rules of all the major platforms relating to, *inter alia*, inciting violence, hate speech, and sharing misinformation (arguments that Kamala Harris offered).

There is also a substantial amount of evidence from press accounts of employee complaints at least at Facebook that high-level executives intervened to protect Trump's social media accounts, along with those of other American far-right figures, against rule-enforcement actions that would otherwise have been undertaken by line employees (Mac and Silverman 2021; Dwoskin, Timberg, and Romm 2020; Frier and Wagner 2020; Solon 2020).

Yet high executive power giveth and high executive power taketh away: After the events of January 6, there was strong reason to believe that the political stability of the United States – which happens to contain the headquarters, the vast majority of the regular employees, and probably most of the assets (unless hidden in offshore tax havens) of the major platform companies – was in severe danger. There was a realistic threat of a coup in the United States; the culpability of the social media companies for facilitating its incitement could potentially have been a fatal public relations disaster if it led to mass user or advertiser defection or a severe legislative response.[21] Accordingly, both Facebook and Twitter finally acted, banning Trump from their platforms as perhaps the most prominent part of the series of company actions that has since been going by the name "the great deplatforming."[22] In both cases, credible media reports suggest that the decision was made directly by top-level

[20] There's no particular reason to think this problem is limited to Facebook, it's just that Facebook is the only company that had such a huge leak.

[21] Moreover, if Trump had actually managed to seize authoritarian rule at that moment, how long would Zuckerberg and Dorsey have been allowed to keep their companies, or their freedom?

[22] Because of its infrastructural role, Amazon's removal of the right-wing social networking company Parler from AWS may have been more controversial within the industry.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

executives, that is, Zuckerberg and Dorsey (Byers 2021) – and Dorsey's very public musing on the decision on Twitter certainly is consistent with this.[23]

There's certainly a story that can be told in which the change in Trump's social media status resulted from good-faith interpretations of platform rules in the face of changing circumstances. For example, the interpretations of those rules might be amenable to change in the context of the broader social threat posed by a user's actions.[24] However, it seems much more likely that decisions around Trump were made in an essentially ad hoc fashion at critical points by company leaders both when they kept his most problematic posts up and when they ultimately banned him.

These events thus illustrate both the benefits and the dangers of top-level executive power. On the benefit side: It has long been understood that a key function of executives in political states is to respond to emergencies, and many scholars have suggested that deviating from or suspending the ordinary operation of law in such states is permissible – or at least inevitable – under such circumstances. Carl Schmitt built an entire theory of sovereignty out of this function of executives (which Chapter 5 will discuss at length – I think the Trump situation actually reveals some of the flaws in Schmitt), and it appears in numerous examples of positive law, such as Article 16 of the French Constitution and the United States National Emergencies Act. In view of the extreme danger on January 6, 2021, both to the companies and to the country in which both companies are headquartered, it is reasonable to defend the actions of Zuckerberg and Dorsey – as well as whichever decision makers at Amazon decided to ban Parler, whoever at Reddit decided to ban various Trump-associated subreddits, and so forth – as necessary emergency steps.

On the other hand, however, there's a plausible case to be made that executive power at both companies brought them – and the United States – to that extremity in the first place. Suppose we accept the – controversial but eminently believable – claims that Donald Trump's social media posts routinely violated Twitter and Facebook rules for years beforehand and that Trump's social media activity was necessary (in a causal sense) to the crisis – that is, that the attack on the Capitol would not have occurred in the absence of Trump's capacity to spread lies and incitement over social media to those of his supporters who were most detached from reality. Then we have to conclude that high-level executives caused the very problem that they were forced to solve at the last moment.

This pathological consequence of unconstrained executive power ought not to be surprising. One well-understood feature of agency is that unconstrained

[23] See the Twitter thread starting at https://twitter.com/jack/status/1349510769268850690 (January 13, 2021) and particularly the reference to "the power an individual or a corporation has over a part of the global public conversation" at https://twitter.com/jack/status/1349510772871766020 – one suspects that one knows who the "individual" is.

[24] For example, from Dorsey's January 13 thread: "Offline harm as a result of online speech is demonstrably real, and what drives our policy and enforcement above all." (https://twitter.com/jack/status/1349510770992640001).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

moment-by-moment decision-making capacity actually undermines the autonomy of an agent, whether individual or organizational (Elster 2000). In particular, the inability to bind one's later self to some constraints radically undermines the capacity to make long-term plans or make commitments that are sufficiently credible to permit the making of deals with third parties. Among other things, this fact is a key justification for the existence of the legal form of contract (Fried 2015, 13–14).

Elsewhere, I have analyzed the problem of costly rule-enforcement by states through this lens (Gowder 2016, 59–62). The analysis directly applies to companies as well. It is likely that company executives saw the enforcement of their rules against Donald Trump before January 6, 2021 as costly in the short term, as they were subject to intense political pressure and threats of retaliation by right-wing political leaders over a supposed bias against conservatives (e.g., Leary and McKinnon 2020). And there is evidence from leaked internal memoranda that at least some Facebook employees understood the problem to be that top-level executives, not constrained by their own rules, were vulnerable to short-term political pressure (Hagey and Horwitz 2021).

So – if the foregoing is true – then at least partial blame for the Trump problem in the first place might be laid at the feet of an inability of platform leaders to bind their own decisions. In other words, if Zuckerberg and Dorsey had, circa 2016 or so, the power to bind themselves to enforce their rules without regard to the identity or political or economic power of the rule violator, they could maybe have controlled Trump's behavior long before it posed a threat. And doing so would also have made the companies more robust against retaliatory threats from the right, since the benefit of those threats to their makers would have been less apparent – people like Josh Hawley would have less reason to believe that the companies would back down in the face of more-or-less empty threats of legislative retaliation – potentially moving those threats off the equilibrium path.[25]

The capacity of commitment to increase an agent's resistance to possibly empty external threats is sufficiently important that it's worth filling out in a little more detail. External estimates of the incentives facing Republican lawmakers during the period when they controlled both Houses of Congress and the Presidency carry a substantial amount of uncertainty. They may have sincerely believed that social media companies were biased against conservatives, or they may have merely been saying that in order to stir up anger in their constituents and prime those constituents to disbelieve things like fact-checking of the misstatements of their political allies on the platforms. Moreover, even if Republican lawmakers sincerely believed that the companies were biased, both threatening legislation and actually seriously

[25] "Off the equilibrium path" is a concept from game theory often used in models of threat and deterrence. Speaking informally, we can understand it in the present context as capturing the idea that a player can sometimes rationally commit to an irrationally costly course of action in order to make it irrational for another player to do the thing triggering the committed-to threat. More, including an example, below.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

attempting to enact legislation come with costs – with the costs of a serious attempt somewhat higher – including the expenditure of political capital in deal-making, possible embarrassment if efforts to legislate fail or if enacted legislation ultimately is struck down by the Supreme Court on First Amendment grounds, and even potential national economic loss from damage done to the companies. Even successful legislative efforts have the cost of no longer being able to use social media "censorship" as a political issue. Against these costs must be weighed the advantages of securing additional opportunities to reach social media users with their messages by deterring rule enforcement against their political team.

Depending on the specific weight legislators give to each of those incentives, there are plausible utility profiles in which a legislator would prefer to threaten to regulate social media firms in order to induce them to grant leeway to their team's content without any intention of actually following through with those threats – particularly if they do not actually believe that such firms are enforcing their rules with a bias against their ideological allies (in which case legislation requiring neutrality would likely be ineffective), or if they do not believe that there is a realistic chance of successfully enacting and enforcing such legislation. Judging by the numerous threats that seem to have gone nowhere even in Republican-controlled branches of government during the Trump administration, it seems likely that one of these utility profiles was in play for leading Republicans during that period.[26] But in the face of uncertainty as to whether all these threats are sincere, it was rational for a company executive to put a thumb on the rule-enforcement scale in favor of the group making the threats, that is, American conservatives, to avoid them being carried out.

Under such circumstances, effectively committing a company to enforcing preexisting platform rules against conservative content would at least partially defang

[26] For example, Missouri Republican Josh Hawley introduced S.1914 in June 2019, which proposed to strip Section 230 protection from companies that engaged in "politically biased" content moderation, but the bill appears to have died in a Republican-controlled committee with no action taken. In September 2020, Mississippi Republican Roger Wicker introduced S.4534, entitled the "Online Freedom and Viewpoint Diversity Act," to substantially limit the scope of Section 230 protection, but that bill appears to have died, like Hawley's, in the Commerce, Science, and Transportation committee – of which Senator Wicker was chairman at the time. In the same year, Georgia Republican Kelly Loeffler introduced both S.4062 ("Stopping Big Tech's Censorship Act"), and S.4828 ("Stop Suppressing Speech Act of 2020"), both of which died in the same committee, as did Lindsey Graham's S.5020 to repeal 230 altogether. By then the Democrats controlled the House, but even when the Republicans controlled both branches in the 115th Congress, there was no action on Texas Representative Louie Gohmert's H.R.7363 ("Biased Algorithm Deterrence Act of 2018"). In October 2020, then-FCC chairman Ajit Pai threatened to engage in a rulemaking process on Section 230, but left office having taken no action (Hollister 2021). It's worth noting by way of caveat that the failure of these efforts to gain traction may not be because they weren't sincere threats – they may have been sincere but withdrawn because they successfully induced enough compliance that the Republicans didn't need to follow through. On the other hand, part of the reason that Zuckerberg and Dorsey felt free to act after January 6 may have been because the lack of follow-through served as evidence that the threats were not sincere (or at least were not imminent, the Democrats having taken control of the White House). Florida and Texas Republicans, evidently under different incentives from their federal colleagues, did manage to legislate at the state level.

the empty threat strategy. There's less reason to make empty threats if their victim cannot surrender to the pressure they create. To be sure, such threats might still be a useful means of voter mobilization (or undermining the credibility of platform fact-checking); however, their incentive and hence their incidence may be reduced to the extent they are motivated at least in part by the capacity to intimidate corporate executives into rule underenforcement.[27]

4.2.1 *Lessons in Self-binding from Political Science*

Political science has a well-developed conceptual apparatus to address these problems of credible commitment (or, depending on the context, sometimes "credible threat"). In the political science context, these ideas appear most prominently in the literature on international relations, relating to the capacity of states to credibly threaten costly military action against one another (e.g., Kilgour and Zagare 1991; Huth 1999), and in the literature on domestic law enforcement, relating to the capacity of states to credibly commit to costly punishment of lawbreakers (e.g., Baker and Miceli 2005). The broad strategic problem is similar across both contexts, so I will simply describe the domestic example, as it is more analogous to the problem faced by platforms.

Consider the following toy problem. A dictator, Caligula, wishes to collect taxes. Naturally, citizens won't pay up voluntarily. Hence, Caligula requires a military/police force to make them do so. However, deploying coercive force is costly – soldiers must eat, ammunition must be acquired, and so forth. Let's suppose that the average cost of punishing a citizen is $1,000. To make life a little easier, we will also assume that the punishment that Caligula can inflict is adequately painful to deter tax evasion, even considering the probability that some tax evaders will not be detected. Unfortunately, Caligula finds that most citizens' tax liability is less than $1,000, and, even if she expropriates all of the assets of every citizen who is found to have evaded their taxes, many citizens' all-in net worth is still less than $1,000. Should someone who is worth less than that amount fail to pay their taxes, it is

[27] A related context may be informative. Network security company Cloudflare terminated the accounts of the Daily Stormer and 8chan because of the vile nature of their content, but ultimately changed their policy to forbid themselves from doing so in large part because exercising such discretion appears to have rendered them vulnerable to external pressure. In the company's words: "In 2017, we terminated the neo-Nazi troll site The Daily Stormer. And in 2019, we terminated the conspiracy theory forum 8chan. In a deeply troubling response, after both terminations we saw a dramatic increase in authoritarian regimes attempting to have us terminate security services for human rights organizations – often citing the language from our own justification back to us." Matthew Prince & Alissa Starzak, "Cloudflare's abuse policies & approach," Cloudflare Blog, August 31, 2022, https://blog.cloudflare.com/cloudflares-abuse-policies-and-approach/. Cloudflare further claimed that "each showing of discretion" in their choices about services to terminate "weakens our argument" in legal challenges to orders seeking to have them carry out global restrictions on, for example, defendants in copyright cases.

irrational for Caligula to expend the costs necessary to punish them: She'll spend more than she can get back. In the absence of some way to commit in advance to punishing everyone who evades taxes, any person worth less than $1,000 will look down the game tree and realize that they're not in any genuine danger of punishment – so they simply won't pay.

Suppose, however, Caligula can make an irrevocable commitment to punishing tax evaders, no matter how expensive it is? The famous doomsday machine in Dr. Strangelove (a movie much beloved by all game theorists) is the paradigm case – an unstoppable machine set to launch a retaliatory nuclear attack without any human intervention. If Caligula can create an unstoppable tax-enforcement machine – even if that machine still costs $1,000 every time it turns itself on – then even a poor citizen will be aware that the machine will come for him if he fails to pay his taxes. Now, every citizen has an incentive to pay their taxes (remembering our earlier assumption that the punishment is painful enough to deter everyone who genuinely faces its threat). And – the most delightful part for Caligula – because everyone pays their taxes, the punishment machine never turns itself on, and hence she never has to pay the cost of punishing – in the lingo of game theory, tax evasion, and its punishment are "off the equilibrium path."

Thus, credible commitments are the canonical way that a state solves its problem of under-punishing. But a state also needs to refrain from over-punishing. "Over-punishing" in this context means using punishment as a means of expropriation, that is, engaging in revenue-seeking punishment in excess of what is permitted by the law. We often use the language of the rule of law to describe the imperative for states to follow their own law, and, at a minimum, it is generally recognized that the rule of law requires the state to only punish citizens in accordance with the law – that is, to refrain from over-punishment (Gowder 2016, 7).

The problem with over-punishment from Caligula's amoral self-interested perspective is that it is widely believed to deter productive economic activity. If my property is not secure against the state – if there is a stated tax rate that is sufficiently low to permit me to profit from investment, but the real tax rate is substantially closer to 100 percent because of the risk of getting looted – then I'm much more likely to attempt to conceal my money or flee the country than to save or invest. And that means a smaller pie for Caligula to tax.

However, once again, there is a problem of short-term incentives to take into account. To see this, imagine again that Caligula is considering whether to punish an alleged tax evader, but, now, the person under her avaricious gaze is quite rich – and quite innocent of tax evasion. Nonetheless, Caligula is powerfully tempted to falsely accuse the rich person of tax evasion and steal all their goods, because, after all, it only costs $1,000 to do so, but the rich person has far more than $1,000 worth of stuff to steal. The time-inconsistency problem arises because Caligula's short-term and long-term interests conflict: If she could credibly commit to not punishing innocent rich people, she could give them an incentive to engage in productive

activity, and hence collect more legitimate taxes in the long run. Thus, according to some scholars, the transition to the rule of law can be explained in part by the desire of leaders to maximize their long-run rents by building institutions permitting them to refrain from short-term expropriation.[28] In political science, Olson (1993; see also Haggard, MacIntyre, and Tiede 2008) argued that leaders of physical territory have an incentive to create functioning legal systems that restrain their own expropriative behavior in order to maximize the rents that they may gain from rule.

Dr. Strangelove's Tax-Evasion Punishment Machine could solve Caligula's over-punishment problem too. To do so, it must control the entire apparatus of punishment, it must only punish those who have failed to pay their ordinary taxes, and its functioning must be known and trusted by the public at large. But, how do we build it?

For platforms, the over-punishment problem is basically the same as the problem for states: In each case, the entity (platform/state) wishes to promote profitable activity (user engagement/capital investment) in the "space" it controls, but, in order to do so, it needs to provide some way to assure those whose activity is required that it won't just totally deprive them of the benefits of their own activity. What prospective influencer will build up a hundred thousand followers and a business based on their content if a slight shift in the political winds inside some company will cause that to all come crumbling down?

With respect to under-punishment, the platform problem is slightly different from the state problem. Platforms, unlike states, probably cannot usually inflict deterrent levels of punishment. At least with respect to social media platforms, with respect to most potential bad actors, the maximum punishment such a platform can inflict (a permanent ban from the platform) is almost certainly not going to be sufficiently painful to deter the worst misbehavior, such as by political propagandists, financial scammers, and the like (who may have teams of fake accounts and reliable ways to optimize on the cost of distributing their lies such that if their content or accounts are removed they will not have lost a too-large investment).

It may be that there is some lingering deterrent effect to the extent that if a platform is particularly effective at eliminating those who engage in rule violations, malicious actors may go looking for softer targets.[29] However, it will be safest to assume that the purpose of platform punishment is, in the classical typology of criminal justice,

[28] Another way to think about this is that Caligula's rate of discounting the future might change – if the regime seems unstable, it might be better to loot the citizenry now; if the regime is more stable it might be better to set up a system to protect long-run economic growth and get a smaller share of a much larger pie over a longer time.

[29] Another exception may be with respect to (a) businesses or politicians that are (b) heavily dependent on a given platform for their revenue or access to voters, and (c) have existing brand/political identities or other goodwill-type assets such that a platform ban is likely to be effective against strategies such as simply creating a new identity and rejoining the platform. Amazon and other transactional platforms might benefit from this kind of deterrent power; so might social media platforms when confronting famous influencers with distinctive individual identities such as real-life celebrities who heavily rely on social media.

incapacitation rather than deterrence – at least in the social media context, platforms need to detect those who are creating a large proportion of the rule-violating content and remove them in order to control the distribution of that content.

Even though platforms lack the capacity to fully bring rule enforcement off the equilibrium path, they still have strong reasons to credibly commit to neutral rule enforcement in order to solve the under-punishment problem:

1. To the extent some marginal deterrent effect is possible, they can remove part of their enforcement costs from the equilibrium path.
2. Credibly committing to neutral rule enforcement may be able to keep some of the political pressure away from platforms. To the extent platforms can point to something akin to Dr. Strangelove's machine and say "See? We aren't making choices about what to enforce!" they are less subject to accusations of bias from both the left and the right.[30] More importantly, if company leaders are unable to succumb to political threats, for example, because someone else controls the rule enforcement system (or checks company control in a robust way), such threats cannot be effective.
3. There may be a marginal effect on user trust from credibly committing to enforcement – a company that makes credible promises to keep hate speech off its platform, for example, may have some competitive advantage, in terms of attracting new users and retaining existing ones, over companies that merely make unenforceable promises.

Because credible commitments solve the over-punishment problem and ameliorate the under-punishment problem, it behooves platform companies to figure out how to make them.

It is important to note that the two functions of credible commitment strategies – to bind an entity to a course of action, and to communicate that binding to external observers in a believable way – are distinct. Commitment strategies are important ways to enforce long-run–oriented behavior even independent of their capacity to signal credibility to outside parties. We might analogize a platform's ignoring or distorting its rules (banning a conservative for political reasons, failing to ban a harassing conservative) to individual health choices such as smoking or eating pizza. Smoking a single cigarette (ignoring a single powerful harasser) might produce more utility than its contribution to long-run pain; however, when this individual rational choice is repeated over an extended period of time, one ends up with lung cancer (an unsafe platform that drives users away).[31] Under such circumstances, some kind of precommitment strategy – that is, some way of making a long-term

[30] Thus, for example, Facebook's efforts to involve third-party fact checkers in its content moderation efforts in order to shield itself from accusations of bias (Lyons 2018).

[31] Cases like these are notoriously problematic from the standpoint of decision theory. I am inclined to see the problem as one of one-off decisions about *di minimis* risks which, when aggregated, are far from *di minimis*, however, this may be an incoherent way to see the problem (Lundgren and

decision to bind the organization to neutrally and completely enforcing its rules in the presence of short-run incentives to the contrary – is advisable.

However, in the platform context, because of the imperative platforms have to maintain user and public trust, a mere commitment to neutral rule enforcement, however ineffective, will be insufficient. Such commitment must in fact be known and believed by (i.e., credible to) outside parties.

4.3 ORGANIZATIONAL TOOLS FOR SELF-BINDING

For individuals, self-binding strategies typically involve recruiting the assistance of external coercion or technology – a set of techniques ranging from Odysseus tying himself to the mast to hear the song of the Sirens (Elster 2000) to applications that allow an individual to increase the cost of undesired behavior by, for example, setting up an automatic donation to one's least favorite politician to punish slip-ups.[32] Organizations, however, have more fine-grained control over their decision-making mechanisms, and can use institutional strategies to shape their own behavior. That is, they may change their own organizational structure in order to change the incentives shaping the entity as a whole.

4.3.1 *Independent Enforcers (Like the Meta Oversight Board?)*

One classic strategy that platforms may borrow from states is to change the identity of the person or entity who implements rule enforcement (including adjudication as a precondition of enforcement), in order to separate the actor who makes a decision about rule enforcement from the actor who feels the pain of the cost.[33]

I have suggested elsewhere that this strategy may play a role in the development of classic rule of law institutions in states, such as the independent judge (Gowder 2016, 59–62). Political leaders may create independent judges or other independent rule-enforcing institutions, and give them incentives to follow pre-existing law, as a precommitment mechanism to enable themselves to engage in costly rule enforcement (prevent under-enforcement). Doing so also protects against over-enforcement to the extent the independent judge doesn't personally receive the benefits of expropriation or is socialized to value legal compliance. Empowering a third-party, in other words, is how we get Dr. Strangelove's punishment machine.

Stefánsson 2020). Other ways of understanding such problems may be in terms of hyperbolic discounting or short-term failures of emotion regulation (Elster 2000, ch. 1). At any rate, the general pattern of such decisions will doubtless be familiar to readers.

[32] For example, Stikk, www.stickk.com/.

[33] Douek (2019, 24–26) draws on the literature on courts in authoritarian regimes to suggest that Meta's Oversight Board can help the company "outsource controversy" by providing a third-party to blame for unpopular decisions. This is an additional benefit of using independent enforcers for credible commitments, but not the most important one.

This idea can apply to platforms as well. A key reason that top-level executives may be tempted to deviate from their own rules is because they are *personally* sensitive to the kinds of threats that might be posed to the company as a whole. Mark Zuckerberg may have been particularly sensitive to the threats of Republicans, causing him to under-enforce Facebook rules against right-wing rule-violating users, because Zuckerberg personally loses a lot of money and status if Josh Hawley successfully takes Facebook's Section 230 exemption away. His employees would be worse off too, but, because, in economic terms, their wealth is vastly more diversified (it's mostly in kinds of human capital that they can convert to cash by working for other companies), they have much less of a felt need to surrender to potentially empty threats. This partly explains why line employees have tended to be stronger advocates for rule enforcement, according to the media accounts cited above, than senior executives: Intuitively, more senior executives are likely to have more firm-specific capital (be less diversified), for example, by having their reputations, networks, and knowledge tied to a specific firm (and perhaps also a higher degree of investment in the firm's stock, as well as compensation packages more tied to the firm's performance).[34]

Unfortunately, judicial independence or anything analogous to it tends to be difficult to achieve and sustain in the world of states because it conflicts with some fundamental imperatives of leadership: Top-level leaders have strong reasons to centralize power in order to maintain their leadership and policy autonomy, and handing over authority to independent enforcers, along with enough sources of power (money, military force) to enforce their own independence undermines that centralization. Thus, trying to use independent enforcers to help leaders constrain themselves may just push the problem back a step: Instead of struggling to commit to costly enforcement of their rules, leaders now struggle to commit to maintaining the independence of their enforcers. This too is a problem for platforms; witness the skepticism about the genuine independence of Meta's Oversight Board in view of the company's control over things like the information it receives and the selection of its initial members (e.g., Newton 2022).

Platform enforcer independence may be easier to achieve than state enforcer independence, if only because platform enforcement does not require the direct application of physical coercion. In the physical world, independent enforcement has to be created by law and backstopped by force, but in the platform world, it can be created by, as Joel Reidenberg (1998) and Larry Lessig (1999) taught us over two decades ago, code. Some degree of platform enforcer independence could be achieved as a purely technical matter, by, that is, engineering direct control over

[34] By way of caveat: The capacity of workers to constrain the companies they work for is limited not only by the relative balance of interests and economic power but also by the ideologies that firms and workers develop to justify what they do. Ari Waldman (2021) illustrates this best in an insightful study of how the concept of "privacy" becomes warped within the workplaces of companies organized around its opposite.

such decisions to personnel within the authority of the enforcer. For example, an independent enforcer for Twitter could have software-level control over the decision to ban or not ban purveyors of disinformation, with the software in question being subject to a third-party audit to ensure that no "back doors" were available to override those decisions.

However, many of the same problems that vex states may still arise in different forms in the context of platform enforcement. There is an inherent trade-off between organizational policy autonomy and the existence of independent enforcers: An independent enforcer, in virtue of its independence, has the capacity to defect from centralized policy decisions. For example, commentators have speculated that Meta's Oversight Board could effectively set aside the company's policy on political advertisements (Levy 2020b).

Moreover, independent enforcers have their own organizational capacity or lack thereof, which may affect the policy/independent enforcement trade-off. Consider, for example, the problem of caseload management: An independent judge who has the capacity to hear many cases (a large staff budget, an efficient adjudication process) can systematically distort policy by defecting from it; an independent judge who only has the capacity to hear a few cases cannot effectively ensure that pre-existing rules are enforced. Hence, a leader trying to empower an independent enforcer still has to make difficult choices even given the ability to use code to entrench its power: Give that enforcer too little organizational capacity, and it may not be able to sufficiently support the kind of credible commitment that the organization needs to make to the outside world; give it too much organizational capacity and it might start imposing its own preferred policies on the broader entity. Put differently, independent enforcement that is effective tends to also entail the delegation of policy autonomy, and there are significant challenges in delegating that policy autonomy in a legitimate fashion; this is, I submit, the foundation of many conventional challenges to constitutional judicial review in modern polities: If we give judges enough institutional capacity to effectively enforce the constitution, we also risk giving them enough institutional capacity to illegitimately impose their own policy choices on elected leaders.

4.3.2 *The Political Foundations of Credible Commitment: Recruiting Workers and Ordinary People to Backstop Self-binding*

In the context of states, many self-control problems are mitigated by democratic institutions that permit mass publics both to exercise some control over policy and to backstop (i.e., by their capacity to sanction political leaders) the independence of judges and other enforcers (Gowder 2014b; Law 2009). This allows for the incentives of policymakers and enforcers to be sufficiently aligned to reduce the risks of enforcement defection, in virtue of the fact that the power of each depends on the willingness of a mass public to support their decisions. For that reason, it is less risky to confer additional organizational capacity on enforcers.

In order to understand the underlying strategic dynamics, please indulge me in a brief digression on how it works for states. Democratic publics suffer from two principal-agent problems: Once they've put an executive in office, that executive suddenly commands a lot of force and engages in a lot of hard-to-observe behavior; how to keep him or her to the will of the public and the laws that (ideally) represent that will in the long run? But, on the other hand, how to keep judges from excessively impeding the pursuit of policy goals by the executive that the people support? The solution is for the judges to control executives, but for this control to go through the threat of collective action by the people, which will only occur when the judges don't defect too badly from the people's present will. That is, in a reasonably well-organized state, the decisions of independent judges who are institutionally committed and socialized to value legal propriety can be used as a signal to trigger coordinated public action (e.g., voting the recalcitrant executive out). So long as those decisions are reasonably well-aligned with the preferences of the public at large, and so long as the public has the capacity to observe executive defiance of judicial officials and engage in collective action, if executives disobey, the public can coordinate on disobedience as a signal to inflict political punishment on executives.[35]

Potentially, platform companies could make use of similar mass-directed policy and enforcement alignment. The "mass" in question could be either (or both) of their employees or their userbases (or even the general public, with some caution about defining that public in an international context and its relationship to a userbase). I will take them up in turn.

Prominent cases of employee activism at many major platform companies suggests that employees have some capacity for collective action – at least during time periods when the technology industry is flush with money for workers (in times of layoffs and contraction, presumably worker power decreases).[36] For example, Google employees organized to prevent the company from doing ethically dubious work for the Pentagon (Wakabayashi and Shane 2018), and Microsoft and Amazon employees extracted at least token concessions from corporate leaders about climate change and work for Immigration and Customs Enforcement (ICE) (Gurley 2019; Frenkel 2018).[37]

Some scholars have suggested that the constraint of state power in historical states has at least in part arisen from the existence of such alternate sources of power within an organization. In premodern states, independent holders of "administrative power" (such as feudal lords), in governments with limited capacity to centralize, have

[35] Obviously, this solution is imperfect, as evidenced by the sorry state of the U.S. Supreme Court right now, which illustrates the difficulty in circumstances of extreme polarization among the public of preventing that polarization from infecting a court; for more details, see Balkin's (2020) account of constitutional rot.

[36] For a discussion of several examples, see Srivastava (2021, 8).

[37] At a smaller software company, a single employee apparently sparked the cancellation of an ICE contract by deleting code that he had written which was being used by the agency (Cox 2019).

extracted concessions from rulers that have led to the constraint of top-level leaders (De Lara, Greif, and Jha 2008). Organized employees may serve as something analogous to the holders of administrative power, to the extent that there is some policy sweet spot that is consistent both with the long-run interests of the company as well as with employee values and interests. If there is such a sweet spot, then employees can demand that corporate leaders follow independent enforcer rulings to the extent those rulings are faithful to that "sweet spot" policy, and hence work together with the independent enforcer to allow corporate leaders to commit to a more long-term oriented rule enforcement strategy. In effect, this becomes a perspective shift on the right-wing critique noted at the beginning of this chapter: Maybe rather than creating short-term defections from company policy, employees can prevent them.

The "administrative power" approach also appears in the literature on contemporary authoritarian governments.[38] A company is a kind of benign authoritarianism as to its internal operations, and some companies are more authoritarian than others, depending on matters such as the extent to which a company is insulated from market controls on its leadership – the example that immediately leaps to mind is Meta, insofar as Mark Zuckerberg's famous stock ownership arrangement (Durkee 2019) effectively guarantees that he can be CEO as long as he wants.

Anne Meng (2020) recently published a monograph beginning with the puzzle of how some authoritarian leaders can transition from a personal regime that cannot survive the death of the leader to a stable authoritarianism (like contemporary China) – a transition that hinges on the building of institutions that stand apart from top-level leaders and can constrain them. But, at a sufficient level of abstraction, this is just the question that we're presented with corporate platforms, and so the same kinds of insights into how it was possible for the Tanzanian National Assembly to constrain Julius Nyerere could shed light on how we might make it possible for the Oversight Board to constrain Mark Zuckerberg.

Meng's answer is that successful institutions empower durable elites other than top-level leaders, which then allows other elites to effectively make alliances (i.e., solve a collective action problem among themselves) sufficiently strong to counteract top-level leaders. In her words:

> When an elite is given a key cabinet position, such as vice president or the minister of defense, he is given access to power and resources that allows him to consolidate his own base of support. Elites who are appointed to positions of authority within the regime then become focal points for other elites. They become obvious potential challengers to the incumbent if she were to renege on promises to distribute rent.[39]

[38] Cf. Douek (2019), who also borrows from the authoritarianism literature.

[39] Meng (2020, 16). Note that Meng isn't developing a theory of third-party enforcement as such. She focuses on the need for authoritarians to build institutions in order to preserve their own rule – in effect, an authoritarian who does so is recruiting allies by, for example, putting someone else in charge

Platforms might also deploy such "elite" kinds of constraint, such as by empowering executives in ethics or oversight roles to constrain those in product and marketing roles who might otherwise undermine platform governance efforts.

Relying on more traditional workplace power management tools, a unionized workplace could write control over rule enforcement into a labor contract, with that control to be enforced by a combination of legal and labor action, or governments could impose regulations providing for intraorganizational insulation of decision-making functions, analogous to regulatory strategies currently used within the financial industry which require separation of functions and of information within organizations and their contractors. For example, the US Sarbanes-Oxley Act prohibits external auditing firms from offering some nonaudit services to the companies that they audit, and the European Union regulates the amount of revenue that auditors may receive from nonaudit services (see Gelter and Gurrea-Martinez 2020, 808–11 for references). Similarly, regulations in many countries require internal boundaries between employees who participate in trading functions and employees who have access to insider information, or effectively require such policies by using their existence as a factor in decisions about insider trading enforcement actions (Dolgoplov 2008; Dahan et al. 2012, 222).

Such an internal separation of powers strategy is readily available to companies and their regulators, for example by requiring the separation of policymaking and enforcing functions from lobbying functions to keep those personnel most susceptible to political pressure away from governance. The same is true about the merger between rule enforcement and customer service noted above: Companies or their regulators may enforce a separation between account managers that service big advertisers or relationship personnel servicing major users and rule-enforcement functions.

A variety of other institutional designs might be available that involve integrating employee decision-making capacity with some other actor's decision-making capacity in order to tune the degree to which a company tracks short-term rather than long-term interests. At the limit, such a strategy amounts to constitutionalizing the operations of a company via workplace democracy.[40] In short, there are many ways that existing company employees, whether junior or senior, could be empowered

of the military or (particularly importantly for Ming) creating a legally designated successor, who then becomes invested in the stability of the overall regime so long as the top-level leader continues providing benefits. By contrast, this chapter focuses on the need for leaders to constrain themselves to make long-term commitments, which they want to do in order to generate stable expectations in others (i.e., trust), which will in turn allow them to draw more benefits – whether that's rents from rule, qua Mancur Olson, or stock value from owning a platform company. At the most abstract level, however, this more or less amounts to the same idea, viz., that top-level leaders can empower and recruit lower level elites to backstop their ability to credibly make promises in the context of shared benefits and cooperation.

[40] On constitutionalizing platforms, generally, see Suzor (2019); on workplace democracy, see Landemore and Ferreras (2016).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

either to backstop an independent enforcer or to serve independent enforcer functions (or other leadership constraint functions) themselves.

However, because technology industry workers do lean to the political left on at least some important issues, institutional innovations that deploy organized employees to prevent short-term corporate failures of self-control may be more effective with respect to the critique from the left than that from the right – that is, handing greater control over content moderation to employees may bring it about that Donald Trump gets banned for threatening violence in violation of Twitter policies, but may not do a lot to prevent conservatives from getting shadowbanned.

4.3.3 *User-Generated Sanctions for Company Commitment*

What about the users and the rest of us? In international relations, one prominent strategy for achieving credible commitment is creating *audience costs*: making leaders vulnerable to sanctions from the general public if they violate their commitments. Fearon (1994; see also Tomz 2007), for example, argued that democratic states have the capacity to buttress the credibility of their public escalations in international military crises, in view of the fact that the leaders of democracies are accountable to domestic audiences which may impose sanctions on them for backing down after vigorous saber-rattling. The "audience" in the theory supposes a third-party to the transaction: The leader of one country threatens the leader of another, with the threat made credible by the external sanctions posed by the first country's citizens. But the idea of making oneself subject to sanctions by some outside party in order to backstop a commitment is more general, and can be applied just as well when the party applying the sanctions overlaps with the party to whom the commitment is directed.

If a sufficiently large group of users to inflict short-term pain on a platform company has the capacity to act collectively, then the company has a short-term incentive to keep from offending them. If companies (or the rest of us) can exercise some influence over the extent to which their users have the capacity to act collectively, to monitor their behavior, and otherwise to effectively inflict sanctions, then the potential exists for institutional design to affect whether platforms can effectively carry out their long-term interest in neutral rule enforcement.[41] This may be particularly appropriate as a strategy under circumstances in which company employees are not trusted, such as with respect to the right-wing critique of social media.

I propose to reinterpret the many existing calls for greater transparency (e.g., Suzor 2019, 136–41; Gillespie 2018a, 198–99; Suzor et al. 2019) in social media content moderation as incomplete suggestions along these lines. Arguments for transparency as a primary solution to the challenges of content moderation make sense if we suppose that external constituencies have some degree of latent power to sanction

[41] Cf. Gowder (2018b), suggesting technological approaches to coordinated consumerism in other contexts.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

platform companies, and merely lack the information to exercise it appropriately.[42] However, transparency solutions cannot actually work if the problem is not a lack of information but the inability to engage in coordinated action to deploy sanctions.

Unfortunately, the status quo makes coordination particularly difficult for platform users. The strong positive network externalities of dominant platform membership mean that it takes a much larger group of users than it otherwise might to credibly threaten to punish a platform. In less jargony terms: The only way users have to sanction platforms right now is by foot-voting (disengaging from, or quitting, platforms). And perhaps people might want to quit a platform in response to its non-neutral rule enforcement, but they might nonetheless benefit too much from being on the platform to do so, unless they can get lots and lots of other people to go with them. Witness the difficulty that many journalists, academics, influencers, and the like are currently experiencing as of this writing (December 2022) in leaving Twitter after the Musk acquisition.[43]

While this effect might undermine the incentive for platform companies to observe their own rules in the short run, it may harm their interests in the long run. A sustained pattern of inconsistent rule enforcement might eventually reach a tipping point at which a company can no longer retain the loyalty of users or its capacity to recruit new ones, at which point the entire ecosystem comes crumbling down and the platform experiences a sudden (but difficult to foresee) mass abandonment.

The nightmare scenario for a platform company would be a kind of abandonment cascade structurally similar to the preference falsification cascades Timur Kuran (1991, 1989) has analyzed. Suppose that different subgroups of users have different levels of tolerance for inconsistent rule enforcement (censorship, failure to get rid of hate speech), where those levels of tolerance also are increasing in the number of users on the platform due to network externalities. Then an abandonment cascade could occur if a platform acts so inconsistently (or appears to do so) as to drive away group A, which (because the degree of positive network externality for

[42] Transparency solutions are challenging for platforms because of the tension between public rules and operational security: Fully public criteria for user behavior and methods for controlling it are unlikely to be sustainable in an environment where sophisticated organized actors (i.e., Russian intelligence agents, among others) are dedicated to subverting platform mechanisms for malicious purposes. This challenge is significant for states as well, of course, as is represented by perennial debates in American law between the paradigms of criminal justice and national security in the war on terror. But at least states have the advantage of a relatively clear-ish distinction between domestic and foreign actors, and secure versus insecure spaces; whereas for platforms everyone in the world is a "citizen" – Russians are perfectly legitimate users of Facebook, and are just as entitled to occupy the same discursive spaces as everyone else, so long as they aren't trying to subvert other societies on it.

[43] In this context, it is at least suggestive that the most prominent effort to quit the major social media platforms has been from the so-called "alt right," a collection of political extremists with a substantial existing alternative media ecosystem that probably made it relatively less costly for them to coordinate to switch from Twitter to Gab, Parler, Truth Social, and so forth – especially since many of their most prominent members had already been chased off the mainstream platform, and hence their presence could not provide an incentive for others to stay.

everyone else shrinks when A is gone) lowers group B's toleration, and hence drives them off; with B gone, C's toleration decreases, and so forth. Such a cascade, as with Kuran's revolutions, could come very suddenly.

There is some evidence to suggest that the cascade model accurately captures platform user incentives: Scholars have identified something like an abandonment cascade in the collapse of Friendster (Garcia, Mavrodiev, and Schweitzer 2013). If this is right, then users may have no effective capacity to collectively (and purposively) threaten sanctions against platforms, but nonetheless pose the real prospect of totally destroying a platform with their aggregate, emergent behavior, out of nowhere, in an abandonment cascade. Platforms might be able to stave off this risk by empowering users to sanction them for trust-betraying behavior before that behavior reaches the point where an abandonment cascade happens – by credibly threatening or inflicting a punishment short of leaving the platform. But merely providing information to users won't do the trick – there must be some noncascade and precascade way of using that information.

This suggests that calls for transparency, at least in the absence of a plausible account of how the general public might actually exercise leverage over companies, are insufficient as a strategy for constraining companies. To be sure, companies have other reasons to operate more transparently in their rule-enforcement. If they discover an effective method of controlling company behavior, transparency about that method is imperative in order to ensure that external stakeholders can actually observe that a company's commitments are, in fact, credible. But the core problem is the structure of the sanctions that might be imposed on a company in the context of the network externalities described above: Platforms are unlikely to have sufficient incentive to keep their users pleased, transparency or no transparency, until the point where an abandonment cascade hits, at which point it's too late. In order to change the shape of this sanction curve, it becomes necessary to provide more intermediate levels of sanction by users or the public at large.

Another way to think about the call for transparency and the broader problem of user-generated sanctions, which might help point the way to more effective techniques, is as an argument for changing the nature of "insider" and "outsider" status with respect to platform conduct. In order for outside actors to constrain companies, some of the knowledge – *but also some of the control* – that had previously differentiated insider versus outsider status – which content is deleted, how the decisions are made – will need to change character. And the company resistance to transparency is partly explained by the need to maintain that status differentiation: By keeping leverage over rule enforcement to insiders, who are vetted by hiring processes and kept loyal by paychecks, platforms ensure that the interests of those who have access to a source of power over their operations are aligned with their interests in areas such as maintaining revenues and protecting against security threats.

But a wide variety of intermediate statuses between full insider and full outsider are possible. Meta's Content Moderation Oversight Board is one example: By recruiting

carefully vetted outsiders who have divergent interests from the company, but whose interests are (hopefully) aligned with the general public rather than with, for example, Russian attackers, and giving them a degree of privileged, insider-level, access to information and authority, Meta does not just propose to give [quasi-]outsiders *knowledge* about what the company is doing, but also *power* over it. If the Oversight Board is to work in the long run, its members must be able to either directly control company behavior (e.g., via technical means), inflict or provoke sanctions on company leaders, or increase the salience and credibility of the existential risk of abandonment cascades if their decisions are defied (and hence make abandonment cascades less likely by giving company leaders a clear message about what behavior is necessary to avoid them).

Another kind of technique to mediate between insider and outsider status has appeared in the corporate law context: the rise of the benefit corporation (Hiller and Shackelford 2018). By making business leaders formally accountable to stakeholders other than their shareholders, such forms of corporate organization have the potential to give those stakeholders some power to exercise constraint over the companies (although corporate law scholars of my acquaintance tell me that this has not been realized in practice) – or, we might say, to radically expand the group of people who are considered "owners," and hence have authority over the organization. More generally, forms of corporate organization like the benefit corporation or other ideas from the broad field of "stakeholder capitalism" (Freeman, Martin, and Parmar 2007; Gadinis and Miazad 2021) might be deployed in order to give outsiders nonexistential threats over platform companies. For example, such tools might be used to backstop neutral platform rule enforcement by giving organized groups of outsiders the legal power to enforce neutrality by filing lawsuits for money damages, the way shareholders in ordinary pure for-profit corporations can (theoretically) file lawsuits against corporate executives who are insufficiently attentive to their duty to maximize profits. Company leaders may have an incentive to confer the capacity for such suits on outsiders as a commitment strategy to publicly force themselves to take the actions most consistent with a company's long-run success.

Here is also where capacity-building efforts by governments may make a particular difference. The rules of contract and of corporate law are controlled by our democratically elected governments, and are malleable. We should consider using them to give platform companies – and others – the tools to subject themselves, in a controlled fashion, to some kind of authority beyond the whims of their leaders. For example, we can give legal teeth to the notion of a benefit corporation, and by doing so confer on companies the tools to in turn confer on the general public some capacity to use the courts to enforce their compliance with public-oriented missions such as providing neutral platforms for speech and sociality. We can create forms of stock ownership that can be conferred on public interest groups to give them some degree of direct leverage over companies. And we can modify labor law to provide

employees with the tools and the incentives to exercise greater voice within companies in order to backstop leaders' commitment to their own public purposes. Some preliminary sketches along at least the last of those lines appear in the conclusion to this book, although as a whole I leave this particular dimension of the approach to corporate governance scholars for further development.

4.4 TOWARD PLATFORM RULE OF LAW

There is a widespread sense that platforms have begun to exercise government-like power, but without the kinds of constraints, such as democracy and the rule of law, that keep government power in check. This idea has been increasingly popular in academia, with, for example, scholars such as Nicolas Suzor (2019) arguing for a new constitutional settlement to bring this power under control, and Rory Van Loo (2021) arguing for extensive procedural protections in their adjudicative processes. Platforms allegedly exercise quasi-governmental power not only in content moderation, but also in copyright enforcement, particularly in the American DMCA "notice and takedown" regime (e.g., Perel and Elkin-Koren 2016), and in business marketplace regulation and dispute resolution, particularly among transactional platforms like Amazon (Van Loo 2016). This sense has also evidently leaked out into the public at large, as evidenced by (thus far unsuccessful) attempts to extend US First Amendment protections (and their state law equivalents) to platforms via formerly obscure doctrines previously extended, at the federal level, only to company towns,[44] and, at the state level, primarily to shopping centers and similar "functional equivalents of the traditional public square."[45] Thus far, such lawsuits have universally failed, but they represent a strong indication that at least some among the public take seriously the notion that platforms are illegitimately exercising quasi-governmental powers.[46] The same is true of legislation enacted in Texas and Florida as of this writing, which purport to prohibit social media political "censorship."[47]

The existence of such legitimacy challenges also creates a compliance challenge. There is well-known empirical research suggesting that compliance with the law depends in part on perceived fairness (Tyler 1990, 1997). From that research, we can predict that to the extent platform rule enforcement is perceived as inconsistent and unfair, users will be less willing to obey those rules.

Accordingly, it will be advantageous for platforms to develop a kind of internal rule of law. By this, I mean systems of constraining their uses of power which follow,

[44] *Marsh v. Alabama*, 326 U.S. 501 (1946).

[45] *Robins v. Pruneyard Shopping Center*, 23 Cal.3d 899 (1979).

[46] See, for example, *Prager University v. Google*, 951 F.3d 991 (9th Cir., 2020) (rejecting First Amendment suit against Google for putting a right-wing nonprofit's YouTube videos in "restricted mode").

[47] As of this writing, both laws are actively winding their way through constitutional challenges with the US Supreme Court as their inevitable destination.

more-or-less loosely, three principles that I have articulated in the context of states (Gowder 2016), adapted for the platform context:

1. Regularity: Platforms should follow their own rules; they should refrain from sanctions or other exercises of authority against users for the platform equivalent of "reasons of state," that is, short-term profit motives.
2. Publicity: Users of platforms should have the opportunity to know the rules that apply to them and to contest their application to their conduct in a procedurally fair way.
3. Generality: Platform rules should be created and applied in a way that recognizes the equal standing of all platform users, regardless of, for example, nationality, gender, race, religion, or political orientation. While this principle does not forbid the differential treatment of some categories of users – for example, the banning of neo-Nazis from social media – it does require that such differential treatment be publicly justifiable in terms that recognize the equality of all to whom the rules are addressed (a criterion that Nazi bans manifestly satisfy).

The implementation of organizational changes to facilitate such a platform rule of law – the creation and support of independent enforcers, and the strategic integration of employee power and outsider scrutiny into rule creation and enforcement processes – has been the topic of this chapter. If legitimacy and compliance scholars such as Tom Tyler are right, such organizational changes may also help solve platforms' broader governance problems by facilitating user compliance.

There is also a moral reason for such an endeavor. We have normative requirements for legal and constitutional institutions, and those requirements may be sensibly applied in part to platforms, generating questions such as:

- Are decision-making institutions really and truly independent of those with interests in the decisions?
- Are there determinate rules that actually bind the platforms (their leaders)?
- Were those rules made in a way accountable to those who are supposed to benefit from the rules?
- Are decisions made in a fair way?

The widespread answer among the public to most of those questions is "no." But these are precisely the issues that the international rule of law development enterprise has concerned itself with (at varying degrees of competence and avoidance of colonialism) in the world of states. In my own work on the rule of law, which attempts to learn from those experiences and from the history of the rule of law, one of the key conclusions that I've drawn is that effective rule of law institutions almost always depend on the threat of collective action by sub-elites and/or the great mass of people to hold the powerful to account.

As applied to the platform context, this suggests that rule enforcement institutions, and the rules themselves, have to be sociologically legitimate. If the Oversight

Board says "Meta has to do X," and we want to deploy market sanctions in order to force Meta to actually do X, then the decisions of the board have to be legitimate enough to motivate people to get mad and seek out alternative platforms if Meta disobeys. It also requires that people have some way of engaging in collective action. If they're actually motivated by forcing a company to stick to its own rules, they need effective tools to act on that motivation, that is, to learn from trusted rule-interpreters when the company has broken the rules, and to coordinate their behavior.

This is, ultimately, the recipe for the credible commitment of a powerful entity to follow its own rules: public, decentralized, power that can be collectively used to hold rule-enforcers to compliance, backstopped by credible and legitimate monitors administering credible and legitimate systems of rules and signaling to the public when the rules have been broken. And this entails a deep integration of ordinary people in governance. For the rules themselves need to be the sorts of things that people want to collectively enforce. The obvious path to this is to make them democratic, that is, to give the people, whether citizens, users, democratic governments, civil society, and so forth, some say in platform rules, rather than having it just be Mark Zuckerberg who determines the conditions under which someone like Donald Trump can or cannot post insurrectionary material on social media. The rule of law development framework can at least give us some criteria for telling whether we've succeeded: We can say that a rulemaking process is democratic in the right way if it forces the kind of alignment between people and interests that can actually draw on collective action for its support. Such democratic institutions can help achieve the platform rule of law insofar as participatory rulemaking has a legitimating function *and* participatory adjudication provides ordinary people with information about one another's interests and beliefs about platform conduct in order to facilitate coordination.

Chapter 6 of this book sketches a preliminary design for some of the institutions that might be put to work to bring this about. But before getting there, we should look at the most developed existing attempt to create an independent enforcer for a platform rule of law to test it against the theoretical material developed thus far. That is the task of Chapter 5.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

5

Actually Existing Platform Self-constraint … Up to a Point

The Meta Oversight Board

As of this writing, the most assertive step toward building an institution potentially capable of meaningfully expanding the capacity for platform governance is Meta's Content Moderation Oversight Board.[1] Conceived in 2018 (Klonick 2020, 2449–50) and consciously modeled on something like the US Supreme Court, the Board issued its first decisions on January 28, 2021.[2]

This chapter considers the model of platform governance which the Board represents in the context of the problems raised by the rest of the book. It is in part a qualified defense of the Board: I argue that an entity like the board can help platforms build short-term responses to emergencies like the January 6, 2021 autogolpe attempt into sustainable long-term rules. However, ultimately, no "Supreme Court"-like entity can solve the problems considered in the previous chapters on its own. Rather, platform adjudicators ought to look less like judges and more like juries – for the knowledge they are required to deploy is not specialized expert knowledge on rules of law but contextual grounded knowledge of the conditions of their local environments and the interaction between platform activities and those local contexts.

[1] The primary sources for the description of the Oversight Board and its purposes and history in this chapter are the following: (1) The charter of the Oversight Board, as posted online at https://scontent-ort2-2.xx.fbcdn.net/v/t39.8562-6/93876939_220059982635652_1245737255406927872_n.pdf as of May 26, 2020 (cited in this chapter as Charter, by section and subsection); (2) the Bylaws and Code of Conduct of the Oversight Board, as posted online at https://scontent-ort2-2.xx.fbcdn.net/v/t39.8562-6/93836051_660280367850128_4544191419119566848_n.pdf as of May 26, 2020 (cited in this chapter as Bylaws or Code of Conduct, by article and section); (3) the June 27, 2019 public consultation report released by Facebook, as posted online at https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-2.pdf (report) and https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf (appendices) as of May 26, 2020. Note that Appendix E to the Consultation Report is the comparative institutions report co-authored by me and Facebook's Director of Product Policy Research, referenced above, for which I was paid – see the appendix to the introduction for details; and (4) an op-ed by the four co-chairs of the Oversight Board, Catalina Botero-Marino, Jamal Greene, Michael W. McConnell, and Helle Thorning-Schmidt, "We Are a New Board Overseeing Facebook. Here's What We'll Decide," *New York Times*, May 6, 2020, www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html.

[2] Oversight Board decisions, www.oversightboard.com/decision/.

The defense of the Oversight Board noted above also entails developing some more conceptual and normative ideas about the notion of platform identity. This chapter sketches a kind of platform legal identity similar to the constitutional patriotism developed by some scholars in the context of states, about which I have written elsewhere (Gowder 2019). I contend that this theoretical work might contribute to addressing some of the underlying controversies associated with the power of platform companies.

5.1 WHAT FUNCTIONS MIGHT THE OVERSIGHT BOARD SERVE? DOES IT DO SO WELL?

In the abstract, we might categorize the functions that an Oversight Board might carry out for Meta and for the outside world into six buckets: (a) propagandistic, (b) informational, (c) corrective, (d) constraining, (e) reformist, and (f) inclusive. I shall ultimately argue that there is a seventh function, which we can call "rationalizing," which the Oversight Board is most likely to serve – and which fits into a broader story about a kind of platform rule of law. But that last one requires rather more theoretical development, whereas the original six buckets are somewhat more conventional.

Propagandistic functions include insulating Meta from external criticism by creating the appearance of oversight and encouraging the perception that decisions are attributable to neutral outsiders. This description is self-consciously neutral as to whether that perception matches reality or not.

While "propaganda" carries negative connotations, not all propagandistic functions are necessarily bad; for example, to the extent Meta is subject to political threats on the basis of false claims of partisan bias in content moderation and to the extent those threats depend for their political force on convincing the public that company personnel are deliberately engaging in political censorship, the propagandistic effect of the Oversight Board may be beneficial for the rule of law-esque reasons discussed in Chapter 4. If the Board is also trusted to make fair decisions, it could improve adherence to those decisions (i.e., reduce efforts to evade content policies or protest and resistance to them) and shield Meta from external political pressures by promoting their broad-based sociological legitimacy. Thus, Klonick (2020, 2426) observes that the Board may be a "convenient scapegoat for controversial content-moderation decisions." But this might be a good thing: If the Board is a scapegoat for *rule-compliant* content-moderation decisions, then that amounts to insulating company executives from paying the political costs of their compliance, and hence facilitating the credible commitments described in Chapter 4.

However, there is also a dark side to the notion of "propaganda" to the extent the Board also insulates Meta or its executives from the pressure they *ought* to experience – more generally, the Board may also amount to what Flew and Gillette (2021, 240) characterize as "pre-emptive self-regulation" which "inhibits the development

of a regulatory framework for platforms as a whole." The capacity of the Board to serve propagandistic functions for good or for ill largely depends on how credible its independence and authority are to outside observers.

The notion of an *informational* function identifies that adjudicators can surface information about governance problems by giving individuals who experience those problems an incentive (in the form of the increased likelihood of having their complaints satisfied) to communicate that information to the public, regulators, and the company itself – with communication to the first two of those mediated by Board decisions (which are public). That potential is present for a wide variety of problems within and without the company – an Oversight Board case could draw attention to some way in which otherwise-reasonable platform rules were causing unintended harm due to novel external circumstances, or to the way in which platform rules and processes themselves are unreasonable on their own terms or underenforced. For an example of how this is already occurring, note that the Oversight Board's adjudications, in conjunction with media reports, have recently drawn attention to troubling features of Meta's internal governance such as the "cross-check" system (discussed in Chapter 4). In doing so, the Board potentially subjects the company to greater public accountability (Schechner 2021).

A *corrective* function is simply the capacity to fix individual incorrect decisions, relative to some standard that includes – but is not necessarily limited to – consistency with platform rules. In view of the relatively low stakes of most individual decisions, this is in some sense the least interesting function of any adjudicator. But sometimes the stakes are high, with the quintessential example being the Donald Trump ban which this chapter considers in detail.

Constraining functions were the subject of Chapter 4. In the context of the present typology, we can understand constraining functions to simply be the aggregation of large-scale informational and corrective functions: That is, by identifying deviations from pre-existing company commitments (in the form of content moderation rules) to internal and external constituents with the capacity to sanction decision makers, and by identifying the commitment-complying (correct application of the rules) decision, an adjudicator can give those constituents the resources to effectively demand that decision makers follow their prior commitments (Hadfield and Weingast 2013). This, of course, depends on its genuine independence for the reasons described earlier.

Reformist functions are closest to those of a stereotypical activist constitutional court, such as the Warren Court in the United States. An adjudicator with sufficient capacity to enforce precedential decisions can directly modify the policies of those whom it regulates by decreeing new rules.[3]

[3] That capacity might come from the ability to decide cases in bulk and hence directly implement those decisions, or from sufficient legitimacy to motivate empowered third parties to enforce those decisions in new contexts.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

Finally, *inclusive* functions capture the overarching democratic aims of this book. An adjudicative body can have a set of decision makers different in morally or practically important senses from those responsible for the underlying decisions which it reviews. Accordingly, it can supply otherwise neglected constituents with an avenue to influence outcomes with respect to any of the other functions. In the context of Chapter 3 of this book, inclusivity can also mean *localism*, that is, incorporating knowledge from those who are closer to the site of some governance challenge.

To some extent, the Board makes improvements in these respects at least from the baseline state of affairs at Facebook/Meta before its creation. The requirement that board panels have a representative "from the region which the content primarily affects" introduces some degree of localism and inclusivity to its decisions.[4] However, in view of the fact that the "regions" are extremely large, in some cases entire continents or bigger, and of course that the board is an elite institution, it is unlikely that significant local knowledge could be incorporated in this fashion.[5]

The financial arrangements for independence (key to the constraining function) are also – at least tentatively – convincing. So far Meta has contributed almost three hundred million dollars to a trust to support the Board's arrangements, giving some reason to believe that (assuming that it doesn't retain control over trust personnel or decisions) the Board will be capable of being reasonably independent at least until the trust money runs out. However, as is always the case with adjudicators operating with limited time horizons, there is reason to worry that Board members may be tempted to shape their decisions to curry favor for future employment opportunities. The fact that Board decisions are unsigned may mitigate this risk.

However, the Board will face several key challenges. First is institutional capacity. The fundamental struggle for all platform content moderation efforts is the sheer volume of cases to be considered, and the Board's likely inability to hear a truly large quantity of cases without (for reasons described below) undermining its independence creates an upper limit on its capacity to review company decisions.

If the Board can only hear a handful of cases relating to particularly salient or policy-relevant (i.e., precedential) conflicts, then it may be able to provide some external

[4] Board Bylaws Art. 1, Sec. 3.1.3; for discussion, see Klonick (2020, 2471). Additional Meta localism exists in its "trusted partner program" by which it seeks input from civil society organizations around the world, however, it is unclear how much actual influence such organizations have. See Meta Transparency Center, "Bringing Local Context to our Global Standards," January 28, 2022 (updated), https://transparency.fb.com/policies/improving/bringing-local-context. Moreover, it's unclear whether the purpose of this program is genuinely to seek input on content rules or simply to comply with hate speech regulations, as an outcome of the negotiation between several platform companies and the European Union described by Bloch-Wehba (2019, 45).

[5] The regions are: "United States and Canada; Latin America and the Caribbean; Europe; Sub-Saharan Africa; Middle East and North Africa; Central and South Asia; and Asia Pacific and Oceania" (Bylaws Art. 1, Sec. 1.4.1). As Douek (2019, 33) points out, a member from one part of a particularly diverse region – such as "Sub-Saharan Africa" or "Asia-Pacific and Oceania" – is unlikely to be all that capable of applying local knowledge to another part.

input into difficult policy decisions, but it is unlikely to be able to control enough outcomes or hear enough complaints to exercise a constraining role or support broad-based inclusion or procedural justice and hence its legitimating capacity will necessarily be limited. In terms of the typology of functions described at the beginning of this chapter, the inability to hear many cases reduces its effectiveness at all of them – the extent to which a Board might convince the public that the company is under control (propaganda), communicate to the public at large as well as to company decision makers problems with content moderation (informational), correct moderation mistakes, reform poor policies, or include diverse voices all scales with the number of the cases it can hear. That is, for each additional case it hears, that's a new chance to exercise public control over the company, learn and teach what's going on in the system, fix a mistake, exercise authority over a policy, or translate the voice of an underrepresented user or stakeholder (or Board member) into an outcome.

Yet as the number of cases the Board chooses to hear grows, its organizational challenges increase: A higher-volume Board may delegate more responsibility to staff, who may exercise undue influence over decisions, undermining its independence. Alternatively, it may need to designate subpanels to render decisions, which may undermine its consistency. The challenge of managing volume has, in other adjudication bodies, led to compromises in the authority of adjudicators for this reason.[6] Douek (2019, 6–7) has suggested – and she's obviously correct – that the problem of volume renders the Oversight Board incapable of providing something like individualized "due process" to users – instead, its function is to serve as a check on the general shape of the rules and their enforcement; yet at the same time, it's unlikely to be wholly effective in shaping company norms partly because of the difficulties of transmitting its results to "the globally distributed and time-starved workforce of content moderators that make the first instance content moderation decisions" and partly just because it lacks the "legitimacy" and "authority" to do so.

That being said, these workload pressures may still permit the Board to exercise the core function of insulating Meta from both internal and external pressure to deviate from its rules in the sorts of particularly high-stakes decisions where such pressures may be most threatening, at least to some degree. There are several preconditions for it to serve this function.

First, it must be genuinely costly for Meta to disobey its rulings; in particular, it must be more costly for Meta to disobey the Board's rulings than for it to obey them, even given the capacity of external actors such as disgruntled politicians to impose sanctions.[7]

[6] See discussion in Gowder and Plumb, "Oversight of Deliberative Decision-making: An Analysis of Public and Private Oversight Models Worldwide," Appendix E to Oversight Board global consultation report, https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf, 162–168; see also Klonick (2020, 2490).

[7] Douek (2019, 47–48) articulates a related, but, I think, mistaken critique. Drawing from an argument of Mark Tushnet's about the ineffectiveness of external checks on the powers of authoritarians, she

It is too early to assess whether such costs are available, but it seems to me that there is reason for concern in this respect, as such costs must be imposed either legally (i.e., through real government judicial sanctions on the basis of Meta's having violated its contract with the Board), or politically/economically (i.e., through public disapproval of company disobedience, and hence public response either through the political process, demanding more direct government regulation, or through the market, by abandoning the platform), and it is unclear that either avenue is readily available in the case of the Board.[8]

Second, workload considerations arise again to raise the concern that the Board may not be able to review enough cases to effectively serve as a constraining check in this sense under some circumstances. For example, if it can only review high-stakes individual cases, but external sources of pressure also care about low-stakes cases in the aggregate (e.g., pressuring the company to under-enforce hate speech rules against large numbers of individuals whose behavior has little individual impact but lots of impact taken together), then Meta may still be vulnerable to pressure in the kinds of cases that it cannot practicably delegate to the Board.

If the Board prioritizes selecting cases in which faithful compliance with company rules is likely to subject the company to costs that cannot so easily be inflicted on the Board itself – for example, cases involving powerful politicians and media figures – then it may be able to enhance its ability to serve a constraining function. Unfortunately, this doesn't address the problem of low-stakes cases that are high-stakes in the aggregate.

As to high-stakes individual cases, the Board ought to prioritize cases in which Meta might have an underlying temptation to break its own rules. That category includes those implicating the interests of external sources of illegitimate pressure, like demagogic politicians. It might also include the review of decisions to leave up content that is likely to be particularly profitable, such as that associated with high-revenue advertisers or popular content producers. For similar reasons, Meta ought to listen to Douek's (2019, 40–41) suggestion to provide the Board with review

argues that the Board "does not actually constitute a 'check' on Facebook's power [when] its actions remain in Facebook's best [long-term] interests." But this is an overly narrow view of what a "check" might be. It may be, and in Chapter 4, I argued that it is the case that the long-term interests of a company – like the long-term interests of Mancur Olson's stationary bandit – are aligned with those of the general public, while the short-term interests of a company are not. Under such circumstances it counts as a perfectly good "check" for the company or the dictator to have institutions that protect it from weakness of will, internal agency problems, and other kinds of pressures leading it toward short term decisions that conflict with the public good, in favor of long-term decisions that support it.

[8] In the case of political/market responses, company compliance would have to be sufficiently visible – either because of mandated company disclosure or some kind of post hoc investigatory powers by the Board. Moreover, the general public (or, perhaps, advertisers, who might have impact individually at a certain size) would have to care enough about compliance to coordinate on sanctions. I am uncertain whether the latter is the case. Moreover, ongoing controversies about things like Facebook's cross-check system, in which a journalist has alleged that Meta lied to the Oversight Board (Horwitz 2021), suggest that the capacity for monitoring compliance is likely insufficient.

authority over algorithmic recommendations and advertising – those are areas in which short-term financial temptations may lead the company to deviate from its rules.[9] Likewise, if programs like cross-check are allowed to continue, it may also be worth considering submitting membership in that program to Board review, as cross-check is essentially a list of people who get special solicitude because they're likely to be able to impose costs on the company.

Democratization, by conferring on ordinary people the capacity to participate in case selection and adjudication, could potentially mitigate the Board's workload-related problems by increasing effective staffing. For example, a multi-tiered system similar to American courts of appeal could be developed with regional popular adjudicators (such as a pool of users chosen from each country) serving as a first layer of appeal from day-to-day content moderators, whose decisions would be subject to appeal to the Oversight Board. Effectively, such a supplemental system could both introduce local knowledge to the adjudication process and serve as a form of workload management by refining issues and filtering meritless cases before reaching the Oversight Board. At the same time, it could create genuine deliberative opportunities closer to the front line of content moderation decisions, and thus both potentially improve uptake of the Board's decisions (since it would be easier to transmit them to intermediate appellate boards than to time-pressured mass workers), and deliver more protective process to individual users.[10]

Currently, the participatory character of the Board is present, but thin. It holds a fourteen-day public comment period for each case that it takes, giving the general public an opportunity to weigh in at will. However, this public comment process is likely to be subject to the standard weaknesses of public comment processes in other high-stakes environments, such as administrative agencies: Those with narrow interests are likely to have a stronger incentive to participate than members of the general public, the Board can largely choose what it does with the comments, and knowledge of the commenting process itself along with the skills to participate effectively are likely to be relatively elite resources. (This is especially so given the short two-week comment period – by the time someone who isn't plugged into

[9] Currently, the Bylaws (Art. 3, Sec. 1.1.2) contemplate future extension of the Board's authority to advertisements. Douek (2019, 42–44) also aptly raises a concern about nonremoval sanctions. It is unclear whether the Board has the authority, or will ever have the practical capacity, to review a variety of other kinds of "soft" sanctions such as reductions in visibility or demonetization. The risk of failing to carry out these expansions of the Board's authority is not merely that various forms of injustice to users or the public might go unreviewed, but that decision makers within the company might have an incentive to use one form of sanction as a substitute for another – to, for example, choose to reduce the distribution of some category of content rather than to take it off the platform – in order to evade Board review, and this might undermine the company's credible commitment to the Board as independent adjudicator.

[10] Of course, it might still be objected that the sheer volume of content moderation would overwhelm those intermediate entities as well. Even if that is true, they could nonetheless deliver *more* individual due process and introduce *more* contextual knowledge. One need not make the perfect the enemy of the good.

the Oversight Board process finds out about a case, it may be too late for them to comment.) Still, the comment process is better than nothing, and given that the Board makes comments public, it can have a potentially beneficial effect insofar as it facilitates scrutiny of the Board's reasoning process – ruling out, for example, clearly inadequate responses to public comments which could impair the Board's reputation.

In terms of the capacity to carry out reformist functions – or, on a more cynical story, to substitute its policy judgments for those of Meta personnel – in addition to institutional capacity issues noted above, there is also some degree of ambiguity as to the capacity of the Oversight Board to generate new rules of platform "law." Its charter provides that while the outcomes of individual cases are binding, the Board's policy guidance to the company is advisory, although formal policy advice will be addressed by the company.[11] Moreover, the charter provides that prior decisions "will have precedential value and should be viewed as highly persuasive" in "substantially similar" cases.[12] It's not terribly clear to me whether this is meant to be read as binding authority, persuasive authority, or something in between – but the Board's own practice may fill that out.[13]

Practically speaking, there are several ways in which Board decisions may have an impact beyond individual cases. First, the mechanisms for company response to policy advice may constrain the company by forcing it to give a reasoned explanation for its policies that can survive public scrutiny (cf. Klonick 2020, 2464). Second, to the extent the Board acquires in the future the institutional capacity to decide a large number of cases, it may constrain the company in practice simply by making rulings based on its own precedent. Third, it may influence company enforcement decisions to the extent those decisions are made "in the shadow" of subsequent rulings by the Board, especially if Meta is likely to suffer a cost from being scolded by the Board in some subsequent case for repeating the mistakes that the Board had already identified.[14]

The extent to which these sources of influence are effective will likely depend on several factors. How much external attention (and hence pressure) can Board decisions generate? How many cases can the Board effectively handle? In other words, if the Board decides to declare a new rule or interpretation, it must either be able to implement that rule/interpretation itself in future cases, or it must have sufficient

[11] Oversight Board Charter, Art. 4, https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf; see also Oversight Board Bylaws Art. 3, Sec. 2.3 www.oversightboard.com/attachment/326581696050456/.

[12] Oversight Board Charter Art. 2 Sec. 2.

[13] Klonick (2020, 2463–64) reports that there was some disagreement in the design process on this question, which potentially explains the resulting ambiguity.

[14] To some extent, a formal capacity to generate precedent would also permit the Board to implicitly expand its institutional capacity, in the sense that decisions which it renders proposing major changes to Meta rules would have a broader effect on other cases.

sociological influence that the threat of scolding from it is meaningful to company decision makers.[15]

As a whole, we must evaluate the Meta Oversight Board as a promising, but limited, source of external constraint on Meta's decisions. The Board will never be as effective as an institution that genuinely empowers ordinary users to intervene on company decisions. However, it is likely to be reasonably effective in cases where stakes are extremely high, in which its decisions are most likely to draw the attention of regulators and the public at large, and in which there is the largest need for an external decision maker to check the unbounded authority of company personnel. In accordance with these suggestions, I will now turn to a direct examination of the Board's highest stakes case thus far – its decision regarding Donald Trump's indefinite suspension from the platform. Below, I argue that this decision demonstrated the Oversight Board's genuine potential.

5.2 A DEFENSE OF THE OVERSIGHT BOARD'S TREATMENT OF THE TRUMP CASE

Midway through the writing of this book, the event that seems to be becoming known as the "great deplatforming" happened – when Donald Trump, toward the end of his presidential term, was evicted from every major social media platform; at the same time, the notorious hard-right "free speech" social media platform Parler was also chased out of the Apple and Google app stores and even its Amazon hosting service. Everyone in the world knows why: The platforms had been used to plot an armed mob attack on the United States Congress aiming to stop the certification of Trump's election loss; Trump himself had made social media posts and speeches inciting that attack.

Trump's removal was also the Oversight Board's first major test, for Facebook's action against his account was submitted for its review. Fortunately, the Board rose to meet the challenge, affirming his removal in a public decision after receiving almost ten thousand comments from the public – but at the same time insisting that the removal not be "indefinite," and demanding a formal reconsideration of Trump's

[15] Thus, we cannot simply suggest, with Schulz (2022, 244), that Board interpretations of Meta rules amount to rule amendments. However, as I have suggested in the past (in a report for Facebook, no less), even adjudicative bodies without formal precedent-setting power tend to develop informal bodies of precedent. See Paul Gowder and Radha Iyenga Plumb, "Oversight of Deliberative Decision-Making: An Analysis of Public and Private Oversight Models Worldwide," report prepared for Facebook in the context of Oversight Board development process, distributed as Appendix E to Facebook, "Global Feedback and Input on the Facebook Oversight Board for Content Decisions," June 27, 2019, https://about.fb.com/wp-content/uploads/2019/06/oversight-board-consultation-report-appendix.pdf, 172–3. It is notable that the Wikipedia ArbCom – the closest prior example to a platform court – seems, according to qualitative research that included conversations with some of its members, to have developed something like an informal system of precedent leading to at least some control over Wikipedia policies (Forte, Larco, and Bruckman 2009, 66).

removal after a time certain in order to reassess the danger posed by his continued access to the platform. To my mind, this illustrates a key function of post hoc rational reexamination of an emergency decision like the one undertaken after the horrifying events of January 6, 2021: The Board took an emergency exercise of (corporate) executive authority and disciplined it – preserving the protective act but subjecting it to an ongoing framework of rational determination under rules going forward. To see the significance of this, it will be helpful to take a small detour into theory.

For the infamous Nazi (yet still influential) legal theorist Carl Schmitt, the sovereign power of exception or of "commissary dictatorship" is a suspension of normal legal institutions necessary to preserve those institutions in the face of an existential threat (Schmitt 2005, 12–13; 2014, 118–19). This problem of emergency power frames Schmittian accounts of the sovereignty of political states. Because states are under an ever-present threat of emergencies that cannot be encompassed within their existing legal structure, sovereignty entails a kind of reserve capacity or prerogative to deviate from the pre-existing legal rules – to "decide on the exception." Typically, this entails the use of coercive force in some way or another – canonical examples include Lincoln suspending the writ of habeas corpus during the Civil War; Charles De Gaulle using the emergency powers granted by Article 16 of the French Constitution in the Algerian War; or the authority granted under Article 4 of the International Covenant on Civil and Political Rights to derogate from the other rights guaranteed by that instrument in states of emergency.

In the platform context, the Schmittian approach bears a striking resemblance to the "shock and exception" dynamic that Mike Ananny and Tarleton Gillespie have identified in an oft-cited conference paper, in which some terrible thing happens or some terrible platform practice is revealed to and criticized by the public (shock), leading to ad hoc exceptions made to solve the immediate problem (or take the heat off the company), but with no stable governance changes (Ananny and Gillespie 2017). We might consider the "great deplatforming" to be just such an example – at least at first.

While Zuckerberg, Dorsey, and their ilk manifestly did the right thing in chasing the insurrectionists and their leader off their communications tools – nobody has a right to speak directly to an armed mob attempting to overthrow a liberal democracy in their name, not unless they can absolutely guarantee that the only thing they'll say is "go home" – the great deplatforming also raised the tension about the idea of content moderation by revealing a quasi-Schmittian character at the heart of the enterprise of platform governance. Even if the leaders of the platforms ultimately agreed with claims, described in Chapter 4, that Trump's conduct had been violating essentially every platform's rules for a long time, until that moment the companies had not seriously acted to clean his pollution off the platform. Choosing that particular moment to chase him off, if understood as an application of those selfsame rules to be distinguished from the previous applications of those rules to keep him on, required an act of interpretation according to which on-platform conduct that

had not fundamentally changed between January 5 and January 7 suddenly assumed a different meaning in light of both Trump's off-platform conduct (in particular, his calls for "strength" and his lawyer's call for "trial by combat" at that rally) and the conduct of his supporters at the Capitol. Even the most conservative reading of that act of interpretation renders it *sui generis*: Only in view of the unique significance of Trump's speech could it be subjected to such an interpretative effort.[16]

Perhaps the key example of the major platforms taking no action (in Facebook's case) or taking much less serious action (in Twitter's case) before January 6 comes from an infamous 2020 tweet and Facebook post in which Trump threatened people who were protesting against police violence with military force. The context was highly conflictual summer 2020 protests over the police killing of George Floyd, protests that ultimately led to further violence, most infamously by one Kyle Rittenhouse (Sullivan 2021). In other words, the country was a powder keg, and the then-President of the United States took to social media to pour the following gasoline onto it:

> These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts. Thank you![17]

Twitter left Trump's looting/shooting post up but placed it behind an interstitial (Seitz 2021). Facebook, well, "raised concerns" and then Zuckerberg himself begged Trump over the telephone to tone it down. Ultimately, Facebook decided to bow to power and keep the post up – a decision reportedly made by Zuckerberg himself (Dwoskin, Timberg, and Romm 2020; Isaac, Kang, and Frenkel 2020; Swan 2020). In other words, Facebook's decision was an exercise of top-level executive power to keep up a post that, on any reasonable interpretation, was a direct threat to shoot protesters.[18] The contrast between the looting/shooting incident and January 6, and

[16] Arguably, the application of the rules changed with the relevant context, that is, with the fact that there was an ongoing violent attack on the US Capitol. This interpretation is supported by the fact that the Oversight Board decision on the Trump suspension attributes that suspension to rules referring to "events" that are violent and to "genuine risk of physical harm or direct threats to public safety." Oversight Board decision in Trump matter, www.oversightboard.com/decision/FB-691QAMHJ. However, as the Board also notes, the suspension was maintained beyond the duration of the attack due to ongoing threats of violence as well as "his continued insistence that Mr. Biden's election was fraudulent" – that is, conditions and rule violations not significantly different in kind from Trump's behavior prior to January 6th, as described for example in the Kamala Harris letter cited in Chapter 4.

[17] Bowden (2020), quoting the Twitter post, but the Facebook post was reportedly the same.

[18] The looting/shooting post seems to me to be much worse than some of Trump's statements during the attack on the Capitol. The latter at least did include a call for the attackers to go home, however insincere and self-undermining – for example, "This was a fraudulent election, but we can't play into the hands of these people. We have to have peace. So go home. We love you. You're very special. You've seen what happens. You see the way others are treated that are so bad and so evil. I know how you feel. But go home and go home in peace," the tail end of one of the statements quoted in the Board decision.

the direct intervention of Mark Zuckerberg in both cases, supports my interpretation of the "great deplatforming" as a kind of suspension of the existing (de jure or de facto) rules to deal with an emergency – possibly a company-threatening emergency, certainly an emergency threatening the overall liberal-democratic order in the United States in which the companies are embedded.[19]

That dynamic motivates my appeal to Carl Schmitt as a way of understanding what the deplatforming revealed. It's difficult to understand anything that happened with Trump (ironically a deeply Schmittian US executive) without the context of an executive-driven decision-making process responding first to external threats from potential regulators and then to much greater external threats associated with January 6. Ultimately, a state of exception was needed, and was declared.

But, as I said, the great deplatforming was undoubtedly necessary. When armed terroristic mobs are storming the Capitol, their communications must be disrupted in order to undermine their capacity to plan future attacks and to undermine the capacity of their leader to command (provoke? inspire?) such attacks. So, in a book like this, which proposes the importation to platforms of organizational strategies from democratic and lawful governments – strategies that are fundamentally anti-Schmittian – the observed need for an act of sovereign heroism stands as a fundamental challenge. Dare we subject the platforms to internal law and to popular control? Could platforms subjected to internal law and popular control have kept Trump away?

5.2.1 *How the Oversight Board's Trump Decision Serves as a Counterexample to Carl Schmitt*

Unexplored by Schmitt's theory is what happens after the executive act declaring the state of exception and responding to the emergency is complete. I contend that there are circumstances according to which emergency executive action might provide feedback to the overall system of rules and support, rather than undermine, something like the rule of law – for both polities and platforms. Beginning with the state context – I would suggest that considering the aftermath of an exercise of emergency power somewhat turns Schmitt on his head. For raw emergency executive power is self-undermining just because, as described in Chapter 4, the rule of law is itself necessary for effective exercise of power. Executives making use of emergency

[19] By the notion of *de facto* rules suspended in the great deplatforming, I mean to suggest that the fact that the major platforms were ignoring – or at least over-charitably interpreting – their own rules by failing to do anything about Donald Trump for years beforehand amounted to a kind of effective law-on-the-ground giving high political leaders a different set of rules. Interpreted generously, this parallel set of rules was made on the grounds of some analogue to "newsworthiness," or the importance of public visibility into the words and actions of their leaders; interpreted cynically it was motivated by the profitable engagement that Trump's behavior generated and the fear that he and his allies would engage in regulatory retaliation otherwise. As noted in Chapter 4, similar informal policies had evidently been applied to powerful politicians in India and Brazil (Purnell and Horwitz 2020; Marantz 2020).

power need some reason to believe that their commands will be carried out, and that belief in turn depends on a broader institutional context in which they can do things like make costly threats credible. *There is no power without the capacity for constraint.* Hence, emergency powers, to be meaningful, require some way of, in effect, regularizing their use, at least after the fact.

I contend that courts and even quasi-courts like the Oversight Board are suited to carrying out such post hoc regularization. With respect to courts – in countries with functional judiciaries, emergency uses of executive powers tend to be subject to challenge after (and sometimes during) the fact. But the core modality of judicial institutions is reason-giving: The thing that makes a court a court, as opposed to some other kind of authority, is that it states general rules and explains why a given use of power over an individual is justified by those rules. Common-law-style courts, that is, courts that generate precedent which itself counts as an authority in future cases, have the further capacity to apply those rules moving forward.

The confrontation between a court and an executive having exercised emergency power is therefore generative. Whether that court upholds or overturns the use of emergency power, a court can attempt to articulate the bounds of that power, and its criteria for application, with reference to the specific facts of the emergency that was presented by the dispute. In doing so, at least sometimes the court can articulate the rules under which similar acts might be permissible in the future. Such an action can, in effect, bring future emergencies of the same form within the system: The next time a closely related threat appears, the executive may not need to suspend the rules to address it, but may be able to follow the rule laid down by the court in the wake of the last emergency. In effect, a court can rationalize and normalize emergency action on a forward-looking basis.

Sometimes, this judicial power can even be self-limiting in a dialogic fashion. For example: One of the most infamous and rightly condemned decisions of the United States Supreme Court is the Korematsu case, in which the Supreme Court upheld the race-based internment of Japanese-Americans during World War II.[20] The case is deeply evil, and it was finally reversed (in dicta) in another evil, much more recent, case, *Trump v. Hawaii*, which upheld Donald Trump's notorious Muslim Ban.[21] Yet Korematsu in part also represents a kind of domestication of emergency power, for while the case permitted the President to carry out the internment, it also was one of the earliest and most important of the articulations of the "strict scrutiny" standard for judging government race discrimination; a standard that later civil rights organizations could use to argue that race-based government action was "presumptively void" (Robinson and Robinson 2005). This is not, of course, a defense of Korematsu – the case was an abomination against justice. But rather, it's a defense of what a scholar like E.P. Thompson (1975, 258–69) or Lon Fuller (1978, 365–81)

[20] *Korematsu v. United States*, 323 U.S. 214 (1944).
[21] *Trump v. Hawaii*, 585 U.S. ___ (2018).

would identify as a kind of valuable normative character of the process of lawlike adjudication: Even a wicked act, when it is filtered through judges or quasi-judges looking to state rules that respond to reasons and make an act compatible with a legal order, can carry within itself the seeds of its own reform.[22]

This is, I contend, part of what happened – at least *in potentia*, depending on the long run response to the Trump decision within Meta – with the Oversight Board's response to the great deplatforming. Zuckerberg's decision was submitted to the Board, and the Board helped regularize it by articulating the principles that justified the executive action and further integrated those principles into the presumptive guidelines of the platform going forward – so that the next time similar threats arise, they can be accommodated without declaring a state of exception. The Trump case thus illustrates how an entity like the Oversight Board can fit into, and help alleviate, the tension between the Schmittian character of company leadership (especially in cases of emergencies) and the value of self-constraint understood as an analogy to the rule of law.

Observe that the Board's decision on the Trump suspension seems to recognize, at least in part, the emergency character of Facebook's act. While there's a certain lack of clarity to the decision, and specifically to the extent to which Facebook appealed to the uniqueness of the situation in justifying the suspension (or was merely motivated by that uniqueness), there are references not merely to Facebook's pre-existing policies on "Dangerous Individuals and Organizations" and incitement, but also to the need to preserve a peaceful transfer of power in the United States, and hence implicitly to the context of the threat to that peaceful transfer occasioned by an attack aimed at preventing the certification of Trump's electoral loss. Moreover, it is quite clear that the Board perceives the particular sanction imposed – an indefinite suspension, as opposed to a time-limited suspension or full-fledged account deletion – as *sui generis*. With respect to that sanction, the Board expresses some sympathy for the exigent circumstances involved, and, while disapproving of the uncertainty created by the indefinite suspension (which grants excessive discretionary power on an ongoing basis), approves of a very similar process: Time-limited suspensions that are renewable, upon the reasoned conclusion that the ongoing risk of incitement continues at the end of the initial suspension. Effectively, this is

[22] In a weaker sense, a bureaucracy tends to generate internal rules, policies, and procedures to implement an executive command. To bureaucratize such a command is to set up structures of authority and rules to generalize it and apply it across the administered domain. For Weber, bureaucracies are forms of "juristic" or legal authority, which share with the law the appeal to general rules to justify their actions (Weber 1946, 299). Translated into the executive power, this distinguishes two kinds of top-level executive commands: The command "go do X to Y" ("go shoot that dissident," "go ban that particular troll from Twitter"), which is a one-off act, and the executive command "go establish and implement a policy of doing X to Ys" ("shoot all dissidents," "ban all trolls"). The latter, even if issued arbitrarily to respond to an emergency, may have a rationalizing function, as at least *future* cases of similar threats will be subject to being addressed under existing rules as opposed to *sui generis* acts of executive power.

a procedural gloss on the indefinite suspension that requires the company to revisit its decision on a periodic basis (but where the period itself appears to be in the company's discretion, or at least not discussed by the Board). This revisiting, however, must be *reasoned* and hence implicitly subject to review and disagreement like any other form of rule-bound action.

This corresponds fairly well to the post-Schmittian framework I have outlined: Even though the Board acknowledges the emergency nature of the suspension and that at least some of the company's rules were derogated from in the process, it both retroactively justifies the basis for the suspension in terms of pre-existing policies (while providing recommendations for the clarification of other policies that may have been applied arbitrarily), and reforms the actual sanction imposed on an ongoing basis to be more compliant with law-like norms without undoing the resolution of the emergency.

Thus, the capacity for Facebook's executives to respond to the emergency was preserved, as was the actual effect of the action: Facebook wasn't ordered to give Trump his account back. In that sense, it held onto the benefits of the Schmittian executive. At the same time, the Board sketched an outline for future responses, not just to Trump but perhaps to individuals in similar positions more generally (consider that the world presently faces parallels to Trump in other nations with massive Facebook user bases, such as Jair Bolsonaro and Narendra Modi). It did so, in effect, by articulating the implications of Facebook's existing commitments, both in its own rules and in its statements about human rights, to such cases.

Moreover, it fills out those commitments in a way that supports the ambiguous claim to precedential power in its founding documents. For a key example, the Board notes that Facebook applied the principles from the Rabat Plan of Action, a standard for considering incitement to hate developed by the United Nations High Commissioner for Human Rights. Although the Board does not, formally speaking, generate precedent that binds the company – just itself, and weakly, as noted above – the fact that the company already applied the criteria from this internationally recognized human rights framework, and the Board explicitly approved of it, suggests that it could be the basis for a kind of informal system of Facebook caselaw, insofar as decision makers within the company sorting out what to do with the next instance of serious incitement are likely to recognize that relying on the Rabat Plan is more likely to ensure that their actions will be upheld.

Critically, the Board's condemnation of "indefinite" suspensions may actually facilitate, rather than restrain, Facebook's capacity to control the behavior of powerful political leaders on its platform. The problem with an indefinite penalty is that it can be revisited at any time – and thus it puts the people with the authority to revisit that penalty in a strikingly weak position with respect to resisting the pressure of powerful political groups. Until the Oversight Board decision, there was no internal basis for the company to say, to an angry Trump-linked pressure group, "no, our policies require that we only reconsider the case at the following date certain [X],

and at that time, you will be required to demonstrate the following things [Y] in order to show that Trump can return to the platform consistent with the safety and human rights interests underlying our rules." The framework offered by the Board – if implemented by Meta and backstopped by some real sanctions for company noncompliance – would provide just such a basis. In response to pressure groups, Zuckerberg or other executives could offer a neutral reason – compliance with the Oversight Board's command to regularize the terms of Trump's suspension – for considering letting him back on the platform only at a certain time and for bounded and relevant reasons. The fact that such decisions when made will be reasoned and subject to further Board review can further support the message that the decision to maintain Trump's suspension was an act of rule-following, not partisan bias.[23]

5.2.2 *Can Platforms Have a Constitutional Identity?*

There's a sense in which this idea of rationalization is latent in Schmitt's conception of sovereignty, at least as transposed to liberal states and, bluntly, denazified. For the root of sovereign power on Schmitt's account is an identified and bounded people on behalf of whom the sovereign acts. This is what leads to Schmitt's (1996) notion of the friend-enemy distinction: A state as a bounded group depends on the notion of "the political," which in turn is understood as the capacity to point outside and say "these are our enemies" by way of contrast. But pluralistic liberal states tend to be ambivalent at the least toward the notion of defining a people with reference to its enemies, especially after the German home of Schmitt's theories showed where they could all too easily lead. To avoid the dangers of such nationalism, contemporary theorists associated with the idea of "constitutional patriotism" such as Habermas (2001) and Jan-Werner Muller (2007a) have suggested that the people can, in essence, be defined in terms of its legal system and the commitments that system represents toward an ongoing enterprise of legal self-definition.[24]

I contend that we can make sense of platforms as having something like a liberal-democratic legal identity in two senses. First, the interests as well as the more aspirational and clearly articulated organizational missions of platform companies require a framework of functional liberal-democratic political states. That gives the companies a reason to defend the boundaries of that politics and to deny the use of

[23] While this book was in production, Meta announced that Trump's account would be restored. Nick Clegg, "Ending Suspension of Trump's Accounts With New Guardrails to Deter Repeat Offenses," Meta Newsroom, January 25, 2023, https://about.fb.com/news/2023/01/trump-facebook-instagram-account-suspension/. The company's announcement is consistent with the framework described in this chapter, in that Meta acknowledged that Trump's suspension was an emergency act but described reforms to its rules and to the sanction imposed on Trump undertaken to regularize the situation in accordance with the Oversight Board's ruling. The announcement also set out (albeit briefly) reasons for restoring Trump to the platform (as a product of the company's evaluation of the ongoing risk) and specific policies for addressing Trump's behavior going forward.

[24] See further discussion in Gowder (2019, 349–54).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

their platforms to those who would destroy the normal liberal-democratic constitutional order. Second, the practical imperatives of successfully operating platforms that are governable in cross-national contexts are, for the reasons described in the rest of this book, dependent on at least something resembling liberal-democratic institutions in a minimal sense internally, that is, processes for the participatory exercise of reason.

Both the ideologies of the employees and the founders of those companies and the practical conditions for their functioning as businesses assume that they're mostly operating in liberal democracies, or at least that their employees can work out of liberal democracies when the angry government officials come to try to force them to regulate content in ways contrary to their own interests. This can be seen, for example, in the First Amendment defenses that the companies rested on in response to legislative efforts to regulate their content moderation practices in Florida and Texas: In order to keep from being forced to host extreme right hate speech that would drive off their users, they need to be able to defend their own editorial independence.

More abstractly, the notion of a many-to-many networked platform only makes sense in the context of an overall view of the social world which in the first instance conceptualizes individuals, qua users, *as* individuals, that is, as capable of deciding for themselves which associations to engage in and of building multiple layers of association representing as well as crossing between different relationships in which they stand with one another (as consumers, as producers, as co-citizens, etc.) – which, in other words, sees the individual as prior to their existing affiliations and sees those affiliations as contingent and mutable. Understood as liberal in this sense, it should be no surprise that the stated ideology of every major platform has trumpeted its commitment to individual freedom and choice (e.g., Adams and Kreiss 2021, 40–57; Halliday 2012).

In the context of the rationalizing function of law-like adjudication, we might also suggest that a platform's identity is partly constituted by its rules and what it does with them. To be sure, as I have emphasized at multiple places in this book, the rules cannot be cleanly distinguished from the overall affordances a platform offers. The definition of the kinds of activity that can be carried out on the platform, either in a positive sense ("here are what the tools on offer are") or a negative sense ("here is what you cannot do") is a fundamental part of the value proposition of such a platform – and thus, in a capitalist environment, also a company's identity. For a more concrete example, part of the way that Facebook and Twitter are different is that Facebook has its famous "real name" policy, which shapes the kinds of interaction that people expect and experience on the platform, and Twitter does not. It's a different kind of (virtual) space, with a different kind of *telos*, setting different kinds of expectations in pursuit of different kinds of goals.

Connecting those ideas to adjudications like the Trump case in the Oversight Board: When rules are articulated in response to emergency executive acts, we

can interpret the series of events as a kind of practical self-learning by a platform, which fills out a vision of its product-userbase-personnel-society nexus, and hence its identity.[25] In turn, the articulation of that identity potentially makes up a normative defense of the initial executive act. In concrete terms: The legalistic or bureaucratic rationalization and domestication of the great deplatforming could provide a retroactive justification of that act, as an interpretation of it in terms of a commitment to preserve the liberal-democratic order in which the platforms are embedded and invested.[26]

Perhaps, however, I ought to offer a more involved defense of the notion of platform identity, for I believe it might serve a broader function that can contribute to the resolution of some of the other problems articulated in this book.

5.3 CAN PLATFORMS HAVE A LIBERAL-DEMOCRATIC IDENTITY?

The best way to begin thinking about platform identity is to start with the notion of "free speech."

One of the classical challenges in liberal-democratic politics is the problem of "tolerating the intolerant." Simplifying a little bit: Liberal societies have a commitment to values like free speech and the marketplace of ideas. But does that commitment extend to those who espouse the denial of those same ideas? Can a society permit its free speech to be used to promote a censorship campaign; more starkly, can a liberal-democratic society with free and open elections permit candidates of a political party with the espoused ideal of overthrowing the democracy and implementing a dictatorship to stand for office?

Political philosophers have struggled with these problems for generations without making a lot of progress. In the meantime, however, the terrain of the problem has expanded from governments to companies, and especially to social media companies. Those companies typically are founded and run with a commitment to ideals of "voice and free expression" (Facebook 2019), or have declared themselves as being "the free speech wing of the free speech party" (Halliday 2012), or have stated

[25] For the philosophers in the audience: This is self-consciously Hegelian, I admit it – but you should have seen that coming back in the introduction when I appealed to Dewey's very Hegelian democratic learning framework. I make no apologies.

[26] On the possibility of retroactive justification, see Gowder (2019). Such a view, incidentally, turns Schmitt on his head, or, perhaps, twists him beyond recognition: In the presence of a functioning system of post hoc rationalization, the distinction between commissary and sovereign dictatorship first disappears, as the acts of the commissary dictator themselves become partly constitutive of the postcrisis normative order; then the state of exception is permitted to gradually disappear, or at least shrink as a liberal legal order encompasses an ever broader set of possible crises; finally "the political" itself dissolves under the pressure of a conception of identity that focuses not on groups of people but on sets of acts. While this book cannot explore the prospect of reading the experience of conducting platform governance back into political theory in order to unsettle or outright overturn conceptions of sovereignty and state like Schmitt's, I observe here that the capacity to do so is yet another reason to bring together the study of politics with the study of platforms.

as a core value "that everyone deserves to have a voice."[27] And this is not merely a value commitment but also an economic one: Because platform economic models are inseparable from chasing positive network externalities, the default strategy is to operate with an expansive conception of the addressable market (userbase); moreover, hosting a truly wide variety of people (with their associated beliefs, interests, and goals) creates a key advantage of large-scale social media platforms in particular, namely, the immense diversity and hence the immense capacity that such platforms have to facilitate niches for people who otherwise would be unable to benefit from the sociality gains to the network.

Even in liberal states which are in principle committed to a kind of neutrality among conceptions of the good (Patten 2012), their governments still frequently themselves adopt controversial positions as a product of public value. A famous statement of this idea in United States law comes from the Supreme Court case of *Rust v. Sullivan*, which upheld, against First Amendment challenge, a government funding scheme for family planning clinics that restricted recipients from offering abortion services using public money. In the pithy words of Chief Justice Rehnquist: "When Congress established a National Endowment for Democracy to encourage other countries to adopt democratic principles, it was not constitutionally required to fund a program to encourage competing lines of political philosophy such as Communism and Fascism."[28] While a debate about the appropriateness of such a statement is possible, the notable point for present purposes is its resemblance to the notion that the United States has a public identity that is inconsistent with support for communism or fascism.[29]

Platform companies tend to be inclined to offer their resources even to advocacy for views that are radically inconsistent with company values: While there are some limitations associated with things like hate speech rules, and there is some reason to think that platform content moderation is moving broadly in the direction of

[27] YouTube, "About YouTube," https://about.youtube/ ("Our mission is to give everyone a voice and show them the world. We believe that everyone deserves to have a voice, and that the world is a better place when we listen, share and build community through our stories").

[28] *Rust v. Sullivan*, 100 U.S. 173, 194 (1991) (internal citation omitted). Of course, to the extent Rehnquist meant to equate abortion with something like fascism, I do not endorse that implication.

[29] In terms of constitutional law, and the theoretical implications of US law for the idea of liberal neutrality, I am radically simplifying matters for the sake of illustration. In reality, the United States is constitutionally obliged to offer some of its resources even to advocacy for communism and fascism, in the form of public forums – areas of government property, such as parks and streetcorners, traditionally held open to all comers. Also, another way – although an oversimplified and problematic way – to think of the problem in liberal states is that liberal neutrality is in terms of *conceptions of the good*, that is, individual values and goals, not neutrality with respect to the political ideologies enabling or threatening the stability of the institutions that make liberalism possible. But, of course, some conceptions of the good entail stability-threatening political commitments (e.g., various theocratic religious views like integralism), and *democratic* states may also be obliged for democratic purposes to tolerate a wider range of attacks on their fundamental modes of political organization. At any rate, the complexities here run extremely deep, but the comparison to states is merely for the purposes of providing a cognitive entry point into the options for platforms, and I don't aim to contribute to the literature on liberal or democratic neutrality/toleration here.

acknowledging broader political commitments and dependencies (e.g., Horwitz and Scheck 2021), as a whole, the baseline view of most platforms seems to be that even, for example, advocacy for political ideologies that might entail the destruction of the platforms themselves is fair game on those same platforms.

Yet the problem of tolerating the intolerant as it's manifested in the political context generates a kind of internal tension that also applies to platforms. Free speech and toleration as core commitments can carry the seeds of their own destruction. On social media, this problem burst into the public eye for the first time in the "gamergate" crisis back in 2014, which provided the world with the spectacle of many prominent women being driven offline by campaigns of extreme gender-based harassment. In other words, the capacity of social media spaces to host feminist gaming voices, and hence the capacity of feminist gamers to engage in free speech, was seriously undermined by the misuse of the free speech of others. For one particularly clear example, one gamergate technique was "doxing": revealing the personal information of their victims publicly, and thus making it possible for criminal third parties to threaten them with violence for their speech.

Part of this paradox is built into the nature of the "free speech" ideal itself. Speech is not simply additive. Some speech can destroy other speech. Gamergate is one example. America's long struggle with electoral campaign finance is another: While the US Supreme Court has repeatedly held that campaign spending is a form of free speech (buying advertisements costs money), critics of the money=speech equation have long pointed out that those rulings permit the wealthy to dominate politics, and thus effectively stifle the voices of smaller and poorer groups. For platforms as for states, so long as one person's speech can be used to capture, undermine, or destroy the conditions necessary for another person to speak freely – whether that's through whipping up the threat of violent retaliation from third parties or just buying up every possible way that anyone else might get access to listeners, readers, audiences – anyone who wishes to run a forum where free speech is genuinely possible for all must curate an ecosystem in which the affordances of speech are genuinely universally available, and speech-destroying-speech is itself excluded.

Because of a growing recognition of this property of free speech, eight years on, the right response to gamergate seems fairly clear to most of us: The companies should have acted much more vigorously, within the limits of their capacity, to protect the victims of this harassment. The leading scholar of online harassment, Danielle Citron (2019), has given a history of the growing recognition of the expressive harm of permitting online harassment. As she explains:

> Cyber harassment destroys victims' ability to interact in ways that are essential to self-governance. Online abuse prevents targeted individuals from realizing their full potential as digital citizens. Victims cannot participate in online networks if they are under assault. Rape threats, defamatory lies, the non-consensual disclosure of nude photos, and technological attacks destroy victims' ability to interact with others. They sever a victim's connections with people engaged in similar pursuits.

> Robust democratic discourse cannot be achieved if cyber harassers drive victims away from it. Victims are unable to engage in public dialogue if they are under assault.[30]

But we have not yet interrogated the underlying normative standpoint from which that new consensus is to be articulated – US free speech norms? Global human rights? Some combination of those things? After all, the conception of what sorts of speech causes harm to the speech interests of others is not measurable by some neutral criterion – we can't simply count the total number of words uttered and conclude that the speech regime in which that number is highest is, therefore, the best (if nothing else, vicious harassers can be *prolific*). It's doubtful that "better for speech" could be measured at all, but if it can, any measure adopted would necessarily make reference to an underlying value to be promoted, because the judgment in question is ineluctably normative.[31] In other words, platforms, like states, have to come to some kind of position on what "free speech" *is* in order to promote it. The very idea is underspecified, and can only be filled out with respect to a thicker set of goals – consistently pursued, an identity.

There is precedent for the notion of a thick organizational normative identity in this sense. Platforms bear at least a substantial (if not complete) resemblance to some of the larger transnational complex organizations that have vexed political theorists since the Middle Ages. Consider the medieval Roman Catholic Church, or the Knights Templar, organizations that transcended international boundaries, yet exercised their own kind of quasi-legal influence over their members and were (much like platforms) perceived as threatening by local rulers partly in virtue of those facts (Levy 2017). Those organizations are key examples of what Jacob Levy (2017), in his canonical study of the phenomenon, calls "intermediate associations," which have traditionally posed a challenge for theories of government centering on states.

The worries about the relationship between such intermediate organizations, states, and people, are remarkably similar: Levy aptly diagnoses a persistent tension between the risk that intermediate organizations will gain power over their members which in turn deprives them of the rights associated with liberal democracies (if you're a conservative, think of university hate speech regulations; if you're a liberal, think of churches with retrograde views on gender roles), and the countervailing risk

[30] Citron (2019, 130).

[31] For example, we might say that a world with more speakers, or more diverse speakers, or more representation of socially subordinated speakers, is better. But that implies an underlying value judgment about which kinds of speech, and from whom, we value. Missing this point is the core error of the US Supreme Court's campaign finance jurisprudence following *Buckley v. Valeo*, 424 U.S. 1 (1976): The notion that restricting campaign finance is a restriction on speech ignores the fact that, among other things, spending on campaign finance is embedded in a competitive market where one speaker can potentially outbid others, and hence that a lack of campaign finance regulation amounts to a *de facto* restriction on the speech of less wealthy interests – a controversial interpretation of the notion of "free speech" that accepts a tradeoff between an increase in quantity of speech for some in exchange for a reduction in the number of speakers.

that excessive state control over such organizations also deprives individuals of the capacity to exercise those freedoms in association (including by waiving them in the pursuit of shared ends). Compare a persistent worry about social media: Does government control over social media content moderation (like the Texas and Florida laws)[32] protect users' freedom (particularly political freedom) or does it undermine it by impairing their ability to choose to participate in communicative ecosystems under known (if restrictive) terms?[33]

Yet an important point that Levy emphasizes about such organizations, which distinguishes them from liberal states, is that they are purposive: They represent shared organizational ends and hence can be expected to generate (and enforce) behavior that varies from what would be predicted under pure liberal neutrality in pursuit of those ends. Universities are dedicated to the production and sharing of knowledge, churches are dedicated to the spiritual well-being of their members, and so forth.

Social media platforms, like other commercial entities, aren't quite as closely connected to an organizational purpose. There wouldn't be much of a point to joining a noncommercial association without some sense of organizational purpose – the reason one joins a church or a university is, arguably, just to participate in that shared end. By contrast, there may not be a shared end in the same sense to, say, Twitter – the company's end may just be to make money, and individuals may join it simply for individual transactional benefits.

Yet the same can also be true for more conventional kinds of intermediate associations – students may join a university just for a credential in the job market, faculty may join a university just for a cushy job with high job security and autonomy – and indeed, universities have been criticized for a long time for putting financial imperatives ahead of their intellectual mission (the reader who teaches at a large American university may simply reflect on the existence of your football team).[34] Moreover, the commercial character of a platform might also give rise to a second-order organizational purpose to the extent that its revenue model is attached to the instantiation of particular values.

[32] Florida: S.B. 7072; Texas: H.B. 20.

[33] The parallel with the Catholic Church in particular (as opposed to other kinds of medieval-era organizations such as the Knights Templar) is striking because a notable fact about the church is that it generated its own ecosystem of intermediate, and partly autonomous, organizations relative to itself – monastic orders, the Society of Jesus, and so forth, much like Facebook or (especially?) Reddit itself spawns distinctive groupings of people. The Catholic Church also developed its own canon law and courts, much like Facebook.

[34] This is also true of historical examples of intermediate associations. For example, Levy offers the Knights Templar as a core example of an intermediate association But a part of the reason for the destruction of the order was that they'd effectively turned into a banking organization rather than an organization devoted to Christian chivalry and whatnot – a character that became rather inconvenient for them when the King of France needed their money and their debtors saw no particular reason to come to their defense (Nicholson 2021, 73, 80). About the medieval Catholic church and the constant battles over slippages from its religious character from the Constitutions of Clarendon to the Reformation, of course, little needs to be said.

Indeed, that is ultimately the source of social media platforms' commitment to free speech. They have not just an ideological commitment to ideas of free speech arising from the libertarianism of their corporate founders, but an economic model that depends on the promotion of diverse communicative content. And promoting diverse content requires a difficult balance between permissiveness and control to permit diverse *people* on the platform. That is, the core interest of the major social media companies is not just "free speech" in the abstract. Rather, it is free speech in concrete terms of a healthy ecosystem in which people are actually capable of connecting with those who share their goals and interests – and that concrete implementation of the ideal of free speech requires defending the ecosystem against the kinds of toxicity that can drive out beneficial diversity. Social media platforms are markedly similar to universities in this sense: While universities ought to be sites for daring intellectual debate, their mission to promote goods like student learning also rules out giving free reign to kinds of communicative interaction that impair that learning mission.[35] Hate and censorship are *both* anathema to such an ecosystem.

In effect, social media companies have economic – as well as ideological – reasons to run their platforms to promote ends similar to those of a sophisticated – not a naive – liberal democracy, and to promote the flourishing of autonomous groups of their users. At the same time, their well-being is also tied in with the success of liberal democracy in the external world. Repressive states rightly view social media as simultaneously a threat (to the extent they provide their citizens with affordances to organize against them, *a la* the Arab Spring) and an opportunity for surveillance and manipulation (e.g., to the extent pernicious propaganda can reach its victims through the platforms, *a la* Myanmar, the Philippines, Trump, and many other examples). So at best such governments attempt to subvert platform content moderation efforts or coerce them to become tools of the regime, at worst to ban them entirely. And being a social media user in a repressive state, unless one happens to be organizing to overthrow it, is likely a bad idea in view of the likelihood that one's government will abuse platforms to surveil and manipulate one.

With respect to the largest social platforms (and here I mostly speak of Meta/Facebook and Alphabet/Google), in some countries, a platform's userbase effectively *is* the public (e.g., Facebook in Myanmar). To the extent that a platform constitutes a significant part of a country's public sphere, promoting sophisticated liberal-democratic interactions on the platform also amounts to promoting them in the public at large, and hence the goals of providing a welcoming and diverse

[35] In the university context, an instructive example is the case of University of Pennsylvania law professor Amy Wax, who misused her free speech rights to make public comments on the alleged intellectual deficiencies of her Black students (Chotiner 2019). Such speech goes well beyond the bounds of academic freedom because it unavoidably indicates to some of her students that they cannot expect equal treatment in a professor's pedagogical function.

communicative environment and of supporting liberal-democratic institutions can align.[36]

I thus propose that we call this combination of value and economic commitments – which may in part be determined by its rulemaking and adjudicative process – a component of a platform's "identity." We might also use the word "mission," although I prefer "identity" not just because it avoids the cognitive sludge associated with corporate propaganda documents like "mission statements" but also because "identity" as opposed to "mission" captures a sense of the present; not of a thing to be aspired to but a thing to be and to preserve. Unlike mission, "identity" also reflects the recognition of the past: The identity of an entity isn't just something decided on as a goal to achieve but also something constituted by an entire history of behavior.

In this sense, the notion of identity can operate bidirectionally in shaping and rationalizing (in the positive sense) a platform's day-to-day decisions when those decisions are controversial or uncertain.[37] Looking forward, the notion of identity can combine a sense of mission and an appeal to consistency, allowing, for example, a significant content moderation or design decision to be guided by reflection on the way that its outcome can be integrated into a platform/ecosystem/firm self-understanding drawing from past decisions, future goals, economic self-interest, and social values.

The attentive reader may have noticed an equivocation in the previous paragraph, helpfully denoted by a couple of slashes: Is a platform identity a property of a firm or of an entire ecosystem? The answer may be both or either: that is, we might imagine a firm having an identity derived from the views and behaviors of its management, employees, shareholders, and so forth; or we might imagine a broader quasi-*demos* composed of a firm in its social context and its users having such an identity. The best characterization of an identity will depend on the behavioral and organizational context: To the extent a broader group of people shape the key decisions and affordances of a platform – for example, if those decisions are rooted in participatory input from a broader userbase – it will make sense to understand a platform's identity as incorporating those users as well.[38] This is simply an implication of the notion that identity is backward-looking, determined not just by someone deciding "this is our identity" but also by reflecting on a pre-existing course of conduct.

Consider how this idea might interact with a concrete example like gamergate. The first beneficial effect of a focus on identity is a shift in the locus of concern from the competing interests of a bunch of external agents – the demands of various

[36] Of course, we must not forget that this could be understood as a pernicious form of colonialism, but for present purposes, I ask the reader to bracket this idea – I'll address it in a moment – and think simply about the notion of platforms coming with their own ideological commitments.

[37] Douek (2019, 49–50) articulates a similar idea, aptly observing that "Facebook cannot escape the need to make a choice about the kind of platform it wants to be."

[38] Since this book ultimately advocates expanding the number of people who get a say in platform company decisions, it also by extension advocates expanding the scope of platform identities. That isn't an accident.

individuals for a microphone or for protection against harassment – that is, "content moderation" – to what we might call "curation" instead, viz., the notion that there is a counterfactual ideal (or at least better) state of affairs on the platform, represented by, for example, a characteristic way that people interact and experience the communicative ecosystem. In terms of concrete responses to particular situations, it's easy to think that 'what do we want our ecosystem to look like' could have generated a swifter and more effective response to gamergate harassment simply by foregrounding the platform-wide interest in not driving away valuable contributors, particularly when those driven away are members of a group underrepresented in their socioeconomic context (female video game players and developers) whom social media could enable to flourish (and who in turn could enrich social media).

More broadly, the concept of a platform's identity reopens the problem of political bias which was a core obsession of Chapter 4 in the form of claims by both the left and right in the United States that platform rule enforcement is biased against them. There's an inherent non-neutrality to platform rules. Consider the prohibition on hate speech: that's a value judgment that certain social media companies have made (albeit at least in part also as a consequence of an economic judgment that their users and advertisers don't want to see that stuff). And it's a value judgment that is inconsistent with certain major political positions, including what might potentially be mainstream positions in certain societies, at least judging by electoral results. Donald Trump said any number of things in public during the 2016 elections that would qualify as hate speech under Facebook's rules, and the notion of platform identity gives companies some reason to wear that fact on their sleeve and shrug off claims of "political bias."

Platform identity is also useful in a second way which responds to the worries raised by Chapter 4: A governor with a clear identity is more predictable to the governed, and more capable of resolving internal conflict, as well as making ongoing decisions going forward about novel situations. In other words, it is more capable of supporting something like a rule of law.

5.3.1 *Toward Participatory Platform Identity*

The notion of platform identity may seem to re-awaken the worries about colonialism associated with efforts to export liberal ideals to other kinds of community. For example, would the notion of a liberal Facebook or a Twitter identity really be a just guide to governance decisions when a company operates in a Muslim country? And are the self-determination interests of the users in that country really to be subjected to a kind of ideology represented by a platform that is itself framed in self-consciously American terms?

There are at least two answers to this objection, one less convincing than the other. The less convincing idea appeals to a kind of transparency ideal represented by markets. To the extent that a company wears its ideology on its sleeve, the existing

government, as well as the public of a state, can identify that ideology and make a relatively free choice in a market context about whether to permit or participate, respectively, in that ideology. As stated, this response is wholly unconvincing, for, we know that market choices like that are not free in as robust a sense as conventional American ideology would have it (e.g., because of the endogeneity of market preferences, see Satz 2010, 180–81), and indeed the incentives offered by platform companies to people in the global South to use their platforms (such as Facebook's "Free Basics") could easily be seen as a *component of* rather than a *response to* colonialism.

A slightly more convincing version of the first response is that the notion of identity may actually serve as an input to some of the questions surrounding the international involvement of platforms, particularly in repressive countries. To the extent liberal democracy is integrated into the identity of a platform, it may be reasonable to just say that its presence in certain countries – particularly those countries in which a company may experience itself as forced to choose between deferring to the decisions of an illiberal government and substituting its own values for those of a local populace – is inappropriate. It might be that operating in such a country – at least without a way of integrating the people of that country into the company decision-making process directly – is incompatible with being a liberal-democratic platform, and hence the company should simply stay out.

But this last point leads us into a much more convincing response. As I said above, the identity of a platform ought not to be seen as merely a product of the values of its shareholders and core workers, as well as the country in which its main business operations are embedded. Rather, the identity of a platform ought to – under sufficiently inclusive institutional design – be interpreted in Deweyian terms, as heavily influenced by the public (or various publics) interacting with that platform. And if this is right, then I think ultimately the colonialism objection is really an objection to an inadequate process of platform identity formation – one that does not accept input from people in the global South, and hence does not fully permit them to integrate their own reactions to and participation on the platform into their own individual and collective autonomy as equal participants in a social world bidirectionally constructing and constructed by the platforms.

In this way, the developing governance proposals of this book can be partly seen as a (partial) mitigation of the problem of platform colonialism: One of the reasons to accept my argument that we must promote the greater influence of people in the global South on platform governance is that until we do so the platforms will not have a fully integrated identity, capable of administering coherent rules rather than simply operating in a system of, as Ananny and Gillespie (2017) said, "shocks and exceptions." And I think that Schmitt had it exactly backward: Rather than understanding the executive and its power to create the exception as derived from the identity of a people (or a public, as Dewey would say), the identity of a platform public is *developed* over time by the participatory development of institutions that eliminate the necessity of exceptions.

Actually, though, I haven't really drawn this idea from Dewey; it's swiped from James Baldwin:

> The black and white confrontation, whether it be hostile, as in the cities and the labor unions, or with the intention of forming a common front and creating the foundations of a new society, as with the students and the radicals, is obviously crucial, containing the shape of the American future and the only potential of a truly valid American identity. No one knows precisely how identities are forged, but it is safe to say that identities are not invented: an identity would seem to be arrived at by the way in which the person faces and uses his experience. It is a long drawn-out and somewhat bewildering and awkward process.[39]

While Baldwin is talking about the difficult process of creating an integrated American identity through mutual recognition and agonistic reconciliation across its racial divide, a structurally similar point – that the identity of a composite entity or a Deweyian public is learned in part through confronting the demands for inclusion by those who have been excluded – holds in the platform context as well. This argument self-consciously refers to my own prior scholarship, drawing on Baldwin and others (Gowder 2019), about what it might mean for a country like the United States to develop an inclusive constitutional identity over time. At bottom, I think the problem for platforms is similar to the problem for American legal institutions: They were founded on exclusionary terms, and those exclusions have led to claims for inclusion: The only way to avoid persistent lawlessness is with a long-term contestatory process in which those claims are listened to and woven into the pre-existing institutional forms.

A more prosaic way to think about the idea of this chapter is that platform governance institutions are in part developed and functional (or not functional) relative to the identities embedded in the companies as well as the affordances of the platform, as those identities are developed by contestation – including contestation over what governance decisions are made and who gets to participate in them.

Global platforms must have a global identity – the problem of colonialism results, in part, from the attempt to impose a local identity associated with the United States on a global public. But publics are constructed in part out of the institutions through which they might act. And so to genuinely construct a global public, companies must be governed through global democratic institutions. Chapter 6 sketches out one model according to which such institutions might be constructed.

[39] Baldwin (2007, 189); for further analysis of this concept of contestatory political identity in this passage, see Gowder (2019, 397–398).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[1] But democracy and governments can come apart in the opposite

[1] 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.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[2] 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.

[2] 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.[3] 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

[3] 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).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[4]

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 Rousseauvian 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.[5] 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.[6] To the extent of their overlap, this book may be seen as, in part,

[4] Dewey (1927, 137).

[5] 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.

[6] 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."[7] 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

[7] 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://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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).[8] 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:

[8] 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.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[9] 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

[9] 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).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[10] 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

[10] 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.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[11]

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)

[11] 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 cliche 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.[12] 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.

[12] I am unavoidably reminded of Robert Nozick's (1974) minimal state "framework for utopia" in the form of subcommunities.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[13] 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.[14]

[13] 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).

[14] 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 anothers' 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.[15] 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.[16] An ideal form of initiation would have a legal framework for these councils (e.g., in the form of the

[15] Back to Aristotle, ultimately, and the idea of ruling and being ruled in turn.

[16] 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.[17] 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;[18] 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

[17] One imagines fashion influencers on the Instagram panel or journalists on the Twitter panel.

[18] 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.[19] 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.[20] 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.[21] 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

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

[20] "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.

[21] 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.[22] 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.[23] 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

[22] Regional councils should have the same provisions for training, staffing, payment, and decisional autonomy as local councils.

[23] 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.[24] 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.[25] While

[24] 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.

[25] 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.[26]

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.

[26] 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.[27]

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

[27] 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/) 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.[28] 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.

[28] Balancing scientific fidelity with decolonial ambitions, councils in the global South could be overrepresented 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 has 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

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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.[29] 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.[30] 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.[31]

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

[29] 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.

[30] On the democratic function of jury nullification, see Gowder (2021, 53–54).

[31] 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.[32]

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

[32] See Chapters 4 and 5 for more on these motivations.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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).[33]

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.[34] 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.[35]

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 multistakeholder 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.

[33] 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.

[34] Several of the chapters in Haggart et al. (2021) generally describe the debates around multistakeholderism on the internet.

[35] See references given by Van der Spuy (2017, 19).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

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).[36]

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.

[36] 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.[37] 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).

[37] 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 antitrust 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).[38] 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

[38] 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.[39] 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.

[39] 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.

Conclusion

How Liberal-Democratic Governments Can Act Now

In order to conclude this volume, I turn back to states, and particularly the United States and European Union, which are the democracies with the most regulatory leverage over the major platforms. Even in the absence of a more comprehensive program like the model described in Chapter 6, there are interventions we, in our democracies, can make over platforms to improve their capacity to govern and mitigate the harms that they inflict on us and on the rest of the world.

Of course, we must take care with any such interventions to minimize their bias in favor of the interests of powerful states and their economies, as well as their ideological and cultural predilections, even while recognizing that they must be sufficiently compatible with those interests and predilections to be realistic as proposals that could actually be enacted. There is an inherent danger of colonialism and imperialism (as discussed in Chapter 2) in regulations originating from powerful countries in the north.

In addition, the United States and European Union must avoid a catalog of familiar dangers of industrial regulation. First, they must avoid creating *regulatory moats* – that is, imposing rules on companies that are so burdensome that larger and more established companies can use their relative capacity to comply with them as a competitive advantage. This is why the model cross-company council system articulated in the Chapter 6 was designed to (a) have its own staffing rather than requiring a thick interface with company personnel, (b) be paid for by a progressive-ish tax and transfer system that imposes higher financial burdens on the richest companies, and (c) be optional for smaller companies, who may choose to participate as it becomes compatible with their competitive position.

Similarly, our governments must avoid what we might call *multistakeholder capture*: the creation of multistakeholder governing bodies that are vulnerable to domination by companies due to the costs of participating in regulatory processes. We can see this as a kind of regulatory moat, or at least a member of the same family, insofar as both failure conditions for regulation go wrong because they miss that many regulations rebound to the advantage of larger, wealthier,

and more entrenched market participants. Multistakeholder capture has been observed in other internet governance contexts, such as ICANN and IANA (see, e.g., Cohen 2019, 230–31).[1]

The final major danger to guard against is *partisan capture*. In both powerful and less powerful countries, politicians have frequently attempted to use regulation or the threat thereof to coerce powerful platforms to act in their own partisan interests. This threat is why I did not list India as a core potential regulator along with the United States and the European Union: while it is a democracy with a huge market share, the Modi regime currently in power has vigorously abused platforms on its own account as well as attempting to bully them with partisan regulation (Horwitz and Purnell 2021; Pahwa 2021). In view of the similar misconduct of US politicians as described in Chapter 4 and its well-known extreme polarization, it may be that the United States should be removed from the list as well – in all honesty, I only left it in there because, as a US citizen, I find myself perhaps a bit biased by hope that our own toxic politics can still be fixable. Moreover, because many of the biggest companies have their core operations in the United States, it has greater regulatory leverage than anyone else. But in view of American political dysfunction, it may be that the European Union is the last credible regulator standing.

INTERVENTIONS ON THE PLATFORM WORKPLACE

One of the implications of Chapter 4 is that direct intervention on the structure of internal company decision-making can actually help companies make and stick to their decisions; a broader conclusion of the entire book is that more diverse and representative decision makers are likely to do a better job at preventing governance externalities. There are several legislative options that could make immediate improvements in company decision-making processes.

First, a well-understood problem within the industry is the phenomenon of the wrong people being put in charge of platform rulemaking and enforcement, like lobbyists (e.g., Wofford 2022) and customer account managers (e.g., Caplan and Gillespie 2020). This one's a relatively easy fix: governments could simply outlaw this sort of dysfunctional corporate organization in the same way that they have passed laws providing for the independence of financial auditors in publicly traded companies. For example, the US government could bar anyone who has ever been a registered lobbyist or supervises a registered lobbyist from exercising

[1] This is why Chapter 6 focused most of its energy on participatory councils to be operated separately from any individual company – such that market entrants can simply plug into an existing system – but which would specifically *not* be the generators of some kind of scheme of universal rules for all companies to obey and in which companies would have input (and hence into which bigger companies are likely to have more input).

decision-making authority over content policy or the design and ranking of recommender algorithms in a platform company.[2]

Similarly, the inclusion of workers in those decisions is likely to both check the untrammeled power of executives and, to the extent workers (particularly workers in developing countries) are more diverse and have superior access to relevant kinds of knowledge, also improve company decision-making. Accordingly, governments (either the US or Europe) could also mandate worker representation in decisions over content policy and the design and ranking of recommender algorithms. There are a number of possibilities for the degree of worker representation, but one possibility could borrow (much like Chapter 6) from American administrative law and require policy changes to be subjected to a kind of notice and comment process, which would permit workers to formally weigh in on the anticipated impact of such changes and require a reasoned response from the executives making those changes. Workers could also be permitted to make policy suggestions that require a reasoned response. Finally, workers could be permitted, perhaps with a supermajority vote, to make the dialogues described above public. It would be critical in this context for "workers" to be defined in a way that includes employees of contractors carrying out company functions, such as content moderators in social media companies.

With respect to US law, one challenge to such interventions would be the First Amendment. Ongoing debate and even litigation about efforts (including by Florida and Texas) to restrict platform content moderation have revolved around the claim that such laws violate company free speech rights, as the design choices for, for example, the feeds of Facebook and Twitter are expressive.[3] To the extent that these arguments are ultimately accepted, the First Amendment may impede regulations relating to content moderation or the decisional process around content moderation (just as the US government would be unlikely to succeed in trying to give newspaper workers a mandatory voice in high-level editorial decisions.) Moreover, precedents such as *Buckley v. Valeo* have suggested both that corporate political activity (focusing on spending, but plausibly implying the inclusion of the activities of lobbyists in policy hierarchies in companies) is protected by the First Amendment and that efforts to protect an overall discursive environment cannot count as compelling interests justifying the restriction of speech.

However, there may be ways around this problem. With respect to the prohibition of lobbyist involvement in content policy, arguably the goal of such legislation

[2] At the limit, I suppose a company's board and CEO would have to supervise both its lobbyists and its rule makers, however, the corporate law world has devised internal controls in other contexts notwithstanding the existence of CEOs, so I assume that there are regulatory options available. For example, executives supervising content policy functions could have some protection against termination without cause to give them some degree of decisional independence even from CEOs.

[3] For discussion, see Bambauer, Rollins, and Yesue (2022); Bhagwat (2021); Goldman and Miers (2021); Langvardt (2021); Rozenshtein (2021); Kosseff (2019).

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

would be as a prophylactic measure to *protect* company free speech rights by insulating them from short-term government pressure. An aggressive constitutional lawyer might thus defend it as a kind of First Amendment paternalism, which ultimately improves the expressive capacity not – as with the speech ecosystem arguments rejected in Buckley – of third parties but of the regulated entities themselves.[4]

Assuming that novel and aggressive First Amendment argument doesn't work, two further solutions suggest themselves. One is simply for the European Union (with substantially less self-defeating free speech absolutism) to do the job. Another is to expand the scope of workplace interventions in ways that are not targeted to speech.

For example, consider a law applying to all companies with multinational operations and more than some large number of customers (say a million, or five million, or whatever). Each company over that size must have an elected workers' council in each country in which they employ more than a certain number of workers (where "workers" for purposes of eligibility to vote, eligibility for election, and determination of which countries are covered includes employees of contractors).[5] Such a workers' council could have the power to demand notice-and-comment style input along the lines described above in any decision which they believe to have high social impact in their home country. Such a regulation would essentially be a general human rights and workplace democracy (e.g., Ferreras and Richmond Mouillot 2017) provision, which would not be limited to company speech or to social impact that relates to the discursive ecosystem. Indeed, it might be more broadly useful in, for example, environmental protection. Accordingly, it would be unlikely to be subject to significant First Amendment challenge.

Regulators can also directly attack the pathologies of the contract worker system of content moderation and similar functions by requiring companies to more closely integrate contract workers with their regular workforces. For example, regulators could specify a maximum number of reporting layers between any contract worker and a manager directly employed by the home company. In addition to the informational benefits of such a proposal, it might also generate moral benefits in terms of promoting respectful treatment of contract workers – at least bringing them closer to main company personnel would force the latter to confront the inequitable ways that contract workers are treated.

INTERVENTIONS ON PLATFORM INFORMATION

One potential barrier to effective participatory governance or external checks on companies is the near-monopoly companies hold over information about their own internal rules, enforcement processes, and the overall scale of problems. And while many

[4] Cf. Elster (2000) on the ways in which constraints protecting against external pressure can promote the overall freedom of an agent.

[5] As discussed in Chapter 2, some countries, such as France, do have versions of a general workers' council system.

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

companies voluntarily publish their rules (or some portion thereof) and/or "transparency reports" and other disclosures, those publications are limited to information consistent with self-perceived (and sometimes short-term) company interests.

Making matters worse, contractual terms used by companies to control competitive advantage in information, such as nondisclosure agreements for employees, potentially impede both the public disclosure of the scale of company problems and cross-pollination of knowledge about the sorts of problems confronted and the techniques of solution between companies, to civil society, and to any novel organizations created in accordance with suggestions like those in Chapter 6. Moreover, these kinds of confidentiality requirements likely bias public disclosure – when confidential information does come out, it tends to be in the form of "leaks" by employees who may be disgruntled or publicity-seeking, and hence may not present a full picture of the underlying circumstances.

Fortunately, these conditions are readily amenable to governmental intervention. As a first pass, courts in countries where company employees work (primarily the United States) should decline to enforce as contrary to public policy nondisclosure agreements relating to information on matters of public concern, and there should be a rebuttable presumption (to reduce workers' legal costs) that disclosures relating to work in the fields of "trust and safety" or "integrity," broadly defined, are on matters of public concern.[6] Second, organizations created to facilitate platform governance, including those similar to the recommendations of this book, should be given the power to compel disclosure (through, for example, subpoenas) of information relevant to their work.

This proposal has the additional advantage of promoting, rather than impeding, market entry – the opposite of a regulatory moat, it would make it harder for established players with robust governance knowledge to use that knowledge to improve their relative market power. However, as a bias check, the reader should also be aware that it would also work to the advantage of emerging civil society organizations composed of workers in these roles, such as the Integrity Institute, with which I am affiliated.[7] It would also be potentially the target of company opposition motivated by their own interests in secrecy.[8]

[6] This includes the work of content moderators, who are also subject to such nondisclosure agreements (Roberts 2019, 73).

[7] I serve on the Integrity Institute's nonprofit board, as well as its Community Advisory Board, and am a fellow of the Institute. Such organizations face challenges originating in company confidentiality rules in building networks of workers who can be useful to the public and to policymakers.

[8] This opposition may not be entirely self-interested. Arguably, eliminating the shroud of secrecy around such work would pose security threats that also expose the general public to harm by permitting bad actors to game the integrity processes (cf. Roberts 2019, 92–94 on that rationale). However, to a substantial extent, the horse has already left that barn given the vast number of leaks from companies like Facebook. In addition, the longstanding record of abuses from organizations which maintain secrecy in their rule-enforcement activities (like essentially every US law enforcement organization) suggests that this is one situation where the tradeoffs likely weigh on the side of transparency.

INTERVENTIONS ON HUMAN RIGHTS LAW

If it's true that platform companies are becoming "Digital Switzerlands" (Eichensehr 2019), then perhaps they ought to be direct subjects of the international order. There's a scholarly literature focusing on the human rights obligations of such companies or proposing that domestic law as well as company policy ought to subject them to human rights standards – Kaye's (2019b) contribution is the most prominent. But why not cut out the middleman?[9]

That somewhat abstract suggestion could be particularly viable in the context of the increased empowerment of workers and the general public described here and in Chapter 6. Social media councils, workplace councils, intermediate popular adjudicative bodies, and the like could make direct reports to international human rights governing bodies, and hence both alert those bodies of emergent threats as well as impede company efforts to conceal their responsibility after the fact. And international human rights bodies could be empowered to order companies to cease certain activities, or even to temporarily cease operation in certain countries during periods of crisis. Think again of the Myanmar genocide.

One question of course about the application of international law to companies is how its rules would be enforced (especially if they're treated conventionally as "nonbinding," see, e.g., Douek 2021, 40) – but this is notoriously a problem with respect to states as well, so difficulty of enforcement alone cannot be an objection to the application of international law. In fact, it may be easier to enforce international law against companies than states since, of course, companies lack armies and territorial rights of sovereignty, while they may be vulnerable to punishments imposed on them via restricting their access to a variety of internationally accessible or managed resources, such as network backbones and namespaces. International orders could also serve as a trigger for domestic action by governments in which companies operate. More aggressively, company executives could be prosecuted in international courts.

Such uses of international human rights law may be most appropriate in the contexts of preventing (or deterring future acts of culpability in) emergencies that require *greater* control of user behavior, such as the Myanmar genocide, election tampering or intimidation, or the events of January 6. This is essentially the opposite of most invocations of human rights law by scholars of platforms (with the notable exception of Wilson and Land (2021)), who tend to follow Kaye (2019b) and focus on freedom of expression, typically in terms of concerns about excessive restriction of user behavior. A more appropriate focus from within human rights law, I contend,

[9] I am no specialist in international law. But at least one scholar who is, Molly Land, has argued that Article 19 of the International Covenant on Civil and Political Rights (covering freedom of expression) by its terms applies to nonstate actors (Land 2013, 443–49). In another article, Land (2019) argues that to the extent state actors delegate regulatory authority to private companies in carrying out speech-related human rights violations, their actions are also reachable by international law as state action. In particular, she argues that systems of intermediary liability like the Digital Millennium Copyright Act and Germany's NetzDG can constitute delegated state authority (Land 2019, 404–8).

is on the responsibility to protect, which would require positive restrictions on platform communication in such emergencies.[10] This is so for two reasons.

First, subjecting platforms to free speech norms is problematic in virtue of the fact that, as Douek (2021, 51–56) explains, such norms are both contested and subject to substantial interpretive variation. I would add, consistent with the discussion in Chapter 2, that imposing parochial interpretations of such norms is objectionably colonial – why *should* US norms about things like the appropriate amount of respect to be shown to the Thai monarchy or the permissibility of visual depictions of the Prophet Mohammed be exported globally?

To flesh out Douek's (quite sound) point in a bit more detail: while freedom of expression is obviously an important human rights value, in virtue of the extremely diverse collection of legitimate interpretations of the notion of freedom of expression across cultures – as well as the fact, illustrated by gamergate and US campaign finance law, that freedom of expression claims can often be made on behalf of speech that actually undermines others' ability to speak – it seems like a poor candidate for a first human rights principle to apply as a universalistic standard to cross-national platforms. The idea of a human right to free expression works in the context of states as the object of human rights law because states have their own political communities with their own cultural traditions around speech – they have a greater capacity to legitimately interpret the universal human rights norm in a way that is particularly adapted to their distinct polity. While I argued in Chapter 5 that platforms can have an identity in a similar sense, the great diversity of cultures present on any of the larger platforms will necessarily make such an identity much thinner and less capable of supporting a clear interpretation of human rights norms surrounding free expression than, for example, the United States with its centuries of First Amendment law or a Muslim country with its distinctive religious norms about matters like the depiction of the Prophet.

Second is the simple fact that platforms have a much greater capacity to impair the right to be protected from great evils such as genocide, election manipulation, and the overthrow of democratic governments than they do to impair the right to free expression. This is a seemingly controversial claim, judging by the immense amounts of popular and scholarly ink spilled on platform "free speech," but a moment's reflection will suggest that it's obviously true. Free speech exists in at least semi-competitive marketplaces where those who are "censored" by one platform can move to numerous other platforms as well as non-platform modalities of communication.[11] The reason that free speech makes sense as a binding norm for states

[10] For a summary of the responsibility to protect in the state context, see McClean (2008). In addition to the responsibility to protect, it may also be appropriate to apply international human rights principles with respect to the rule of law which can backstop some of the constraint described in Chapter 4.

[11] I don't deny that platform "censorship" could impair a user's expressive interests to *some* degree – I simply observe that the capacity of such "censorship" to so impair is limited by the presence of even less-good alternatives.

rather than platforms is because states have soldiers and police who can shut off *all* methods of expression for a person by locking them in a cell or shooting them.

By contrast, if a platform fails to comply with the responsibility to protect, it can lead to direct and inescapable harm. A person killed because genocidal propaganda was spread on a platform cannot exit that platform to escape the harm. A country whose democracy is toppled because foreign agents spread misinformation on one platform is not somehow saved because some other platform does a better job. Competition doesn't alleviate the dangers to which the responsibility to protect is directed. Only law can help.

INTERVENTIONS ON COMPETITION POLICY

In view of the problem of platform colonialism, governments ought to consider forms of competition policy that are compatible with market entry from non-US countries, particularly those in the Global South. For reasons described in Chapter 2, I am skeptical of policies that make it too easy for dangerous companies such as the proliferation of right-wing extremist social media platforms in the United States to enter the market; however, entry in markets currently underserved by existing systems of platform governance may be relatively free from this worry – if the big platform companies are neglecting the safety of users in some country anyway, we might as well make it easier for competitors from that country who might be more capable of engaging in governance to enter. And although such market entrants remain subject to the related worry that smaller and more local platforms may be more susceptible to bullying by authoritarian governments, it is far from obvious that larger platforms are immune from such bullying (witness Facebook's relationship with the Republican party), especially given that effective governance in a country requires *some* local presence in that country, and any personnel (or empowered users) in a country are points of leverage for governmental coercion.[12]

Grewal (2008, 173–79) has argued for alterations to network structure that permit entry and interaction as a potential remedy to unjust kinds of network power. This idea seems to be helpful in the case of platform colonialism, as well as corresponding in part to policy initiatives that are already partly in place. For example, many advocates have argued for data portability and interoperability as a form of competition policy, and this is partly implemented in the European General Data Protection Regulation (Engels 2016). In principle, it might be possible to implement such regulations in the United States, or more robust regulations in Europe, in a way that favors competitors from countries with lower GDP per capita, in minority languages, and the like. For example, laws requiring interoperability or API access to company data may explicitly limit the entities to which access must be granted in

[12] On the susceptibility of local company personnel to government coercion, whether authoritarian-lawless or democratic-lawful, see for example, Haynes (2016), Horwitz and Purnell (2021).

favor of smaller companies primarily serving users in the Global South. However, I make this suggestion with some hesitation, for I lack expertise in competition policy sufficient to have any confidence about its viability. But in the context of Grewal's analysis of network power, such policy innovations would amount to conferring a greater degree of "compatibility" insofar as the resources of the platforms would be (partly) accessible using alternative corporate (and governance) entities.[13]

Bonina et al. (2021, 892) suggest that platforms in the Global South tend to have a greater degree of involvement by "NGOs, public institutions, governments, and developmental organizations." They also identify the presence of "platform cooperatives … which rely on democratic decision-making and a shared ownership of the platform by workers and users" in the Global South (Bonina et al. 2021).[14] To the extent this is correct (no evidence is cited for the claim), this suggests that promoting platform competition from the Global South has a greater capacity to promote community empowerment, at least to the extent the NGOs and the like are run by people from the countries in question rather than from wealthy countries – an issue which warrants further investigation.

The reference to NGOs and platform cooperatives is a suitable place to close this book. At bottom, such innovations potentially represent a method of operating platforms that is separate from the imperatives of capitalism. In the absence of such imperatives, there is a substantially weakened incentive to create many of the features of our contemporary large platforms that have posed so many governance problems in the first place. For example, nonprofit platforms may have much less reason to build recommender algorithms that prioritize engagement and the drive for boundless growth even when that means serving toxic or inadequately protected markets.[15]

In effect, the recommendations of this book may be seen through the lens of a kind of postcapitalist democracy. Rather than proposing the creation of new, public (and hence non-capitalistic) platforms, I have argued for recognizing the public character of the platforms we already have. They may nominally be private businesses operated for profit, but they carry out functions similar to those of states and exercise – with inexcusable clumsiness – immense amounts of power not only over their users but over the rest of the world.

Thus, the recommendations in this book ultimately amount to taking their quasi-public character and building quasi-public institutions to match it. However,

[13] Similarly, the multilevel system of councils described in Chapter 6 promotes what Grewal calls "malleability" insofar as it would introduce the capacity to change platforms in response to the needs of diverse communities of users and stakeholders.

[14] As I read them, it is unclear whether cooperatives are relatively over represented in those countries.

[15] Even in nonprofit platforms those incentives might not be completely absent – as discussed in Chapter 1, users have limited time to devote to a multiplicity of platforms, so those platforms which do not grow enough to leverage *some* degree of network effects or otherwise attract users to spend their time there might just die off, even if they are not operated for the revenue they generate.

I reject previous commentary which proposes to do so by transposing ideas from older forms of public-private hybridity such as applying the First Amendment to social media companies or treating platforms as (highly regulated and subject to open access requirements) public utilities like electrical company monopolies. Neither of these are appropriate options. The First Amendment is not an appropriate option because, as discussed above, social media companies operate across cultures with wildly different free expression norms, and, as discussed in Chapter 5, because to some extent they have their own distinctive voices attached to their identities as intermediate purposive organizations. The notion of a regulated public utility makes sense as a response to natural monopolies in essential goods and services, but regardless of what one thinks about the anticompetitive character of the platforms, the core governance problems they generate are not, as I argued in Chapter 2, due to a lack of competition. Moreover, we don't want governments to directly be regulating the platforms too extensively, because, also as I argued in Chapter 2, governments often are doing so for the wrong reasons, such as in support of programs of political repression.

Instead, I propose to build something new, albeit rooted in the study of earlier kinds of governance: Directly democratic public-private governance structures meant to be genuinely inclusive and to be tailored, not to the problems of capitalism's past, but to the problems of capitalism's present and future. We are unlikely to return to a world without vast global platforms with diverse groups of users coming into conflict and generating endless novelty. But I believe we can build institutions that permit us, collectively, to adapt to the challenges this scale and diversity present, and hopefully retain the vast gains in interconnection, creativity, commerce, and free expression that those platforms allow while mitigating the threats to public safety and democracy they present. Let's get building!

References

Abrams, Lawrence. 2018. "Google Agrees to Pay $11 Million to Owners of Suspended AdSense Accounts." *BleepingComputer* (blog). May 18, 2018. www.bleepingcomputer.com/news/google/google-agrees-to-pay-11-million-to-owners-of-suspended-adsense-accounts/.

Achen, Christopher, H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Adams, Julia. 1994. "Trading States, Trading Places: The Role of Patrimonialism in Early Modern Dutch Development." *Comparative Studies in Society and History* 36 (2): 319–55. https://doi.org/10.1017/S0010417500019071.

Adams, Kristen, and Daniel Kreiss. 2021. *Power in Ideas: A Case-Based Argument for Taking Ideas Seriously in Political Communication*. Cambridge: Cambridge University Press.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *The Quarterly Journal of Economics* 130 (4): 1885–939. https://doi.org/10.1093/qje/qjv030.

Agarwal, Bina. 2000. "Conceptualising Environmental Collective Action: Why Gender Matters." *Cambridge Journal of Economics* 24 (3): 283–310. https://doi.org/10.1093/cje/24.3.283.

2001. "Participatory Exclusions, Community Forestry, and Gender: An Analysis for South Asia and a Conceptual Framework." *World Development* 29 (10): 1623–48. https://doi.org/10.1016/S0305-750X(01)00066-3.

Aguiar, Luis, Christian Peukert, Maximilian Schäfer, and Hannes Ullrich. 2022. "Facebook Shadow Profiles." arXiv. https://doi.org/10.48550/arXiv.2202.04131.

Alba, Davey. 2018. "How Duterte Used Facebook to Fuel the Philippine Drug War." *BuzzFeed News*, September 4, 2018. www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war.

Alexander, Leigh. 2016. "Facebook's Censorship of Aboriginal Bodies Raises Troubling Ideas of 'Decency'." *The Guardian*, March 23, 2016, sec. Technology. www.theguardian.com/technology/2016/mar/23/facebook-censorship-topless-aboriginal-women.

Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. "Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes." *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 1–30. https://doi.org/10.1145/3359301.

Aligica, Paul D., and Vlad Tarko. 2012. "Polycentricity: From Polanyi to Ostrom, and Beyond." *Governance* 25 (2): 237–62. https://doi.org/10.1111/j.1468-0491.2011.01550.x.

Amazon.com. 2020. "Amazon Establishes New Counterfeit Crimes Unit." *Amazon News*, June 24, 2020. https://blog.aboutamazon.com/company-news/amazon-establishes-new-counterfeit-crime-unit.

Amnesty International. 2022. "The Social Atrocity: Meta and the Right to Remedy for the Rohingya." ASA 16/5933/2022. Amnesty International. www.amnesty.org/en/documents/ASA16/5933/2022/en/.

Ananny, Mike, and Tarleton Gillespie. 2017. "Public Platforms: Beyond the Cycle of Shocks and Exceptions." http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/anannyGillespie-publicPlatforms-oii-submittedSept8.pdf.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme* 3 (1–2): 8–22.

2015. "Moral Bias and Corrective Practices: A Pragmatist Perspective." *Proceedings and Addresses of the American Philosophical Association* 89: 21–47.

Anderson, J. Jonas. 2015. "Court Competition for Patent Cases." *University of Pennsylvania Law Review* 163 (3): 631–98.

Angwin, Julia, Madeleine Varner, and Ariana Tobin. 2017. "Facebook Enabled Advertisers to Reach 'Jew Haters'." *ProPublica*, September 14, 2017. www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters.

Arab News. 2022. "TikTok Sets Up Safety Advisory Council for MENAT Region," February 8, 2022. www.arabnews.com/node/2020911/media.

Article 19. 2018. "Self-Regulation and 'Hate Speech' on Social Media Platforms." Article 19.

2019. "The Social Media Councils: Consultation Paper." Article 19.

2021. "Social Media Councils One Piece in the Puzzle of Content Moderation." Article 19.

Arun, Chinmayi. 2022. "Facebook's Faces." *Harvard Law Review Forum* 135: 236–64.

Bachmann, Jan, and Jana Hönke. 2010. "'Peace and Security' as Counterterrorism? The Political Effects of Liberal Interventions in Kenya." *African Affairs* 109 (434): 97–114. https://doi.org/10.1093/afraf/adp069.

Bain, Marc. 2019. "Nike Is Ending Its Two-Year Fling with Amazon." *Quartz*. November 13, 2019. https://qz.com/1747795/why-nike-will-stop-selling-on-amazon/.

Baiocchi, Gianpaolo. 2003. "Participation, Activism, and Politics: The Porto Alegre Experiment." In *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*, edited by Archon Fung and Erik Olin Wright, 45–76. The real utopias project, volume 4. London: Verso.

Baker, Matthew J., and Thomas J. Miceli. 2005. "Credible Criminal Enforcement." *European Journal of Law and Economics* 20 (1): 5–15. https://doi.org/10.1007/s10657-005-1011-3.

Baldwin, James. 2007. *No Name in the Street*. New York: Vintage.

Balkin, Jack M. 2020. *The Cycles of Constitutional Time*. New Haven: Yale University Press.

2018. "Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation." *U.C. Davis Law Review* 51 (3): 1149–210.

Bambauer, Jane, James Rollins, and Vincent Yesue. 2022. "Platforms: The First Amendment Misfits Symposium: Compelled Speech: The Cutting Edge of First Amendment Jurisprudence." *Indiana Law Journal* 97 (3): 1047–70.

Barnidge Jr, Robert P. 2018. "Anti-Zionism, Jus Cogens, and International Law: The Case of the Banjul Charter and Arab Charter." *The Journal of the Middle East and Africa* 9 (1): 77–92. https://doi.org/10.1080/21520844.2018.1449796.

Baronian, Laurent. 2020. "Digital Platforms and the Nature of the Firm." *Journal of Economic Issues (Taylor & Francis Ltd)* 54 (1): 214–32. https://doi.org/10.1080/00213624.2020.1720588.

Bartz, Diane. 2021. "U.S. Senator Wants to Ban Big Tech from Buying Anything Ever Again." Reuters, April 12, 2021, sec. Technology. www.reuters.com/technology/us-senator-wants-ban-big-tech-buying-anything-ever-again-2021-04-12/.

BBC News. 2018a. "The Country Where Facebook Posts Whipped Up Hate," September 12, 2018, sec. BBC Trending. www.bbc.com/news/blogs-trending-45449938.

2018b. "Facebook Admits It Was Used to 'Incite Offline Violence' in Myanmar," November 6, 2018. www.bbc.com/news/world-asia-46105934.
2020. "Myanmar Rohingya: What You Need to Know about the Crisis," January 23, 2020, sec. Asia. www.bbc.com/news/world-asia-41566561.
2021. "Bolsonaro: New Bill Will Limit Tech Giants' Power to Remove Content," September 6, 2021, sec. Latin America & Caribbean. www.bbc.com/news/world-latin-america-58470093.
Beaubien, Jason. 2018. "'I Would Rather Die Than Go Back': Rohingya Refugees Settle into Life in Bangladesh." *NPR*, August 24, 2018, sec. World. www.npr.org/2018/08/24/641239849/i-would-rather-die-than-go-back-rohingya-refugees-settle-into-life-in-bangladesh.
Bedell, Zoe, and Benjamin Wittes. 2016. "Tweeting Terrorists, Part I: Don't Look Now but a Lot of Terrorist Groups Are Using Twitter." *Lawfare* (blog). February 14, 2016. www.lawfareblog.com/tweeting-terrorists-part-i-dont-look-now-lot-terrorist-groups-are-using-twitter.
Beer, Stafford. 2002. "What Is Cybernetics?" *Kybernetes* 31 (2): 209–19. https://doi.org/10.1108/03684920210417283.
Beitz, Charles R. 2018. "Property and Time." *Journal of Political Philosophy* 26 (4): 419–40. https://doi.org/10.1111/jopp.12145.
Bekos, Paschalis, Panagiotis Papadopoulos, Evangelos P. Markatos, and Nicolas Kourtellis. 2022. "The Hitchhiker's Guide to Facebook Web Tracking with Invisible Pixels and Click IDs." arXiv. https://doi.org/10.48550/arXiv.2208.00710.
Bell, Stephen, and Andrew Hindmoor. 2012. *Rethinking Governance: The Centrality of the State in Modern Society*. Cambridge: Cambridge University Press.
Belli, Luca. 2015. "A Heterostakeholder Cooperation for Sustainable Internet Policymaking." *Internet Policy Review* 4 (2). https://doi.org/10.14763/2015.2.364.
Bellman, Steven, Ziad H. S. Abdelmoety, Jamie Murphy, Shruthi Arismendez, and Duane Varan. 2018. "Brand Safety: The Effects of Controversial Video Content on Pre-Roll Advertising." *Heliyon* 4 (12): e01041. https://doi.org/10.1016/j.heliyon.2018.e01041.
Beningi, Matthew C., Kenneth Joseph, and Kathleen M. Carley. 2019. "Bot-Ivistm: Assessing Information Manipulation in Social Media Using Network Analytics." In Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining, edited by Nitin Agarwal, Nima Dokoohaki, and Serpil Tokdemir, 19–42. Lecture Notes in Social Networks. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-94105-9.
Benkler, Yochai, Rob Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.
Berger, Jonah, and Katherine L. Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49 (2): 192–205. https://doi.org/10.1509/jmr.10.0353.
Berkes, Fikret. 2002. "Cross-Scale Institutional Linkages: Perspectives from the Bottom Up." In *The Drama of the Commons*, edited by Elinor Ostrom, Thomas Dietz, Nives Dolšak, Paul C. Stern, Susan Stonich, and Elke U. Weber, 293–321. Washington, DC: National Academy Press.
Bevir, Mark. 2013. "Governance as Theory, Practice, and Dilemma." In *The SAGE Handbook of Governance*, edited by Mark Bevir, Paperback ed., 1–16. Los Angeles, Calif.: SAGE.
Bevir, Mark, and R. A. W. Rhodes. 2015. *The State as Cultural Practice*. Oxford: Oxford University Press.
Bhagwat, Ashutosh. 2021. "Do Platforms Have Editorial Rights?" *Journal of Free Speech Law* 1 (Symposium Issue): 97–138.

Bhattacharjee, Yudhijit. 2011. "How a Remote Town in Romania Has Become Cybercrime Central." *Wired*, January 31, 2011. www.wired.com/2011/01/ff-hackerville-romania/.

Bickert, Monika. 2018. "Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process." *Facebook Newsroom* (Blog). April 24, 2018. https://about.fb.com/news/2018/04/comprehensive-community-standards/.

Biddle, Sam. 2021. "Revealed: Facebook's Secret Blacklist of 'Dangerous Individuals and Organizations'." *The Intercept* (Blog). October 12, 2021. https://theintercept.com/2021/10/12/facebook-secret-blacklist-dangerous/.

Birnbaum, Emily. 2021. "Facebook Staff Complained for Years about Their Lobbyists' Power." *Politico*, October 25, 2021. www.politico.com/news/2021/10/25/facebook-fatal-flaw-technologists-lobbyists-516927.

Bisbee, James, Megan Brown, Angela Lai, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2022. "Election Fraud, YouTube, and Public Perception of the Legitimacy of President Biden." *Journal of Online Trust and Safety* 1 (3). https://doi.org/10.54501/jots.v1i3.60.

Blattman, Christopher, Alexandra C. Hartman, and Robert A. Blair. 2014. "How to Promote Order and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior through Community Education." *American Political Science Review* 108 (1): 100–120. https://doi.org/10.1017/S0003055413000543.

Bleakley, Paul. 2021. "Panic, Pizza and Mainstreaming the Alt-Right: A Social Media Analysis of Pizzagate and the Rise of the QAnon Conspiracy." *Current Sociology*. https://doi.org/10.1177/00113921211034896.

Bloch-Wehba, Hannah. 2019. "Global Platform Governance: Private Power in the Shadow of the State." *Southern Methodist University Law Review* 72 (1): 27–80.

Bloom, Joshua, and Waldo E. Martin, Jr. 2016. *Black against Empire: The History and Politics of the Black Panther Party*. 2nd ed. Oakland, California: University of California Press.

Bonina, Carla, Kari Koskinen, Ben Eaton, and Annabelle Gawer. 2021. "Digital Platforms for Development: Foundations and Research Agenda." *Information Systems Journal* 31 (6): 869–902. https://doi.org/10.1111/isj.12326.

Bowden, John. 2020. "Zuckerberg Expressed Concern to Trump over Rhetoric amid Protests: Axios." Text. *The Hill* (blog). May 31, 2020. https://thehill.com/policy/technology/500391-zuckerberg-expressed-concern-to-trump-over-rhetoric-amid-protests-axios/.

Bowers, John, and Jonathan Zittrain. 2020. "Answering Impossible Questions: Content Governance in an Age of Disinformation." *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-005.

Bowman, Sam. 2020. "Amazon's Tightrope: Balancing Innovation and Competition on Amazon's Marketplace." *Truth on the Market* (blog). April 27, 2020. https://truthonthemarket.com/2020/04/27/amazons-tightrope-balancing-innovation-and-competition-on-amazons-marketplace/.

Börzel, Tanja A., and Vera van Hüllen. 2014. "State-Building and the European Union's Fight against Corruption in the Southern Caucasus: Why Legitimacy Matters: EU-State-Building in the Southern Caucasus." *Governance* 27 (4): 613–34. https://doi.org/10.1111/gove.12068.

Brady, William J., Ana P. Gantman, and Jay J. Van Bavel. 2019. "Attentional Capture Helps Explain Why Moral and Emotional Content Go Viral." *Journal of Experimental Psychology: General* 149 (4): 746–56. https://doi.org/10.1037/xge0000673.

Braun, Joshua A., and Jessica L. Eklund. 2019. "Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism." *Digital Journalism* 7 (1): 1–21. https://doi.org/10.1080/21670811.2018.1556314.

Brennan, Jason. 2016. *Against Democracy*. Princeton: Princeton University Press.
Brignall, Miles. 2016. "Banned by Amazon for Returning Faulty Goods." *The Guardian*, March 18, 2016, sec. Money. www.theguardian.com/money/2016/mar/18/banned-by-amazon-returning-faulty-goods-blocked-credit-balance.
Brodkin, Jon. 2021. "Amazon 'Seized and Destroyed' 2 Million Counterfeit Products in 2020." *Ars Technica*, May 10, 2021. https://arstechnica.com/tech-policy/2021/05/amazon-seized-and-destroyed-2-million-counterfeit-products-in-2020/.
Broockman, David E., Gregory Ferenstein, and Neil Malhotra. 2019. "Predispositions and the Political Behavior of American Economic Elites: Evidence from Technology Entrepreneurs." *American Journal of Political Science* 63 (1): 212–33. https://doi.org/10.1111/ajps.12408.
Bruckman, Amy. 2022. *Should, You Believe Wikipedia? Online Communities and the Construction of Knowledge*. Cambridge, United Kingdom; New York, NY: Cambridge University Press.
Brynjolfsson, Erik, Avinash Collis, and Felix Eggers. 2019. "Using Massive Online Choice Experiments to Measure Changes in Well-Being." *Proceedings of the National Academy of Sciences* 116 (15): 7250–55. https://doi.org/10.1073/pnas.1815663116.
Bueger, Christian, Timothy Edmunds, and Robert McCabe. 2020. "Into the Sea: Capacity-Building Innovations and the Maritime Security Challenge." *Third World Quarterly* 41 (2): 228–46. https://doi.org/10.1080/01436597.2019.1660632.
Buerger, Cathy. 2021. "Speech as a Driver of Intergroup Violence: A Literature Review." *Dangerous Speech Project*. https://dangerousspeech.org/speech-as-a-driver-of-intergroup-violence-a-literature-review/.
Byers, Dylan. 2021. "'This Is Not Normal': Behind the Decisions at Facebook and Twitter to Deplatform Trump." *NBC News*, January 14, 2021. www.nbcnews.com/tech/tech-news/how-facebook-twitter-decided-take-down-trump-s-accounts-n1254317.
Calo, Ryan, Ivan Evtimov, Earlence Fernandes, Tadayoshi Kohno, and David O'Hair. 2018. "Is Tricking a Robot Hacking?" In. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=3150530.
Campanella, Thomas J. 2011. "Jane Jacobs and the Death and Life of American Planning." *Places Journal*, April. https://doi.org/10.22269/110425.
Canegallo, Kristie. 2019. "Meet the Teams Keeping Our Corner of the Internet Safer." *Google Blog (blog)*. February 5, 2019. https://blog.google/around-the-globe/google-europe/meet-teams-keeping-our-corner-internet-safer/.
Caney, Simon. 2001. "Review Article: International Distributive Justice." *Political Studies* 49 (5): 974–97. https://doi.org/10.1111/1467-9248.00351.
Caplan, Robyn, and danah boyd. 2018. "Isomorphism through Algorithms: Institutional Dependencies in the Case of Facebook." *Big Data & Society* 5 (1): 1–12. https://doi.org/10.1177/2053951718757253.
Caplan, Robyn, and Tarleton Gillespie. 2020. "Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy." *Social Media + Society* 6 (2): 1–13. https://doi.org/10.1177/2056305120936636.
Carley, Kathleen M. 2020. "Social Cybersecurity: An Emerging Science." *Computational and Mathematical Organization Theory* 26 (4): 365–81. https://doi.org/10.1007/s10588-020-09322-9.
Carr, Madeline. 2015. "Power Plays in Global Internet Governance." *Millennium: Journal of International Studies* 43 (2): 640–59. https://doi.org/10.1177/0305829814562655.
Chant, Tim De. 2021. "Apple Sued for Terminating Account with $25,000 Worth of Apps and Videos." *Ars Technica* (blog). April 23, 2021. https://arstechnica.com/tech-policy/2021/04/apple-faces-class-action-lawsuit-over-its-definition-of-the-word-buy/.

Chen, Adrian. 2014. "The Laborers Who Keep Dick Pics and Beheadings out of Your Facebook Feed." *Wired*, October 23, 2014. www.wired.com/2014/10/content-moderation/.

Chotiner, Isaac. 2019. "A Penn Law Professor Wants to Make America White Again." *New Yorker*, August 23, 2019. www.newyorker.com/news/q-and-a/a-penn-law-professor-wants-to-make-america-white-again.

Citron, Danielle Keats. 2018. "Extremist Speech, Compelled Conformity, and Censorship Creep." *Notre Dame Law Review* 93 (3): 1035–72.

2019. "Restricting Speech to Protect It." In *Free Speech in the Digital Age*, edited by Susan J. Brison and Katharine Gelber, 122–36. New York, NY: Oxford University Press. https://doi.org/10.1093/oso/9780190883591.003.0008.

Cohen, Julie E. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. New York, NY: Oxford University Press.

Cole, Samantha. 2021. "'WHYYYYY?': Police Upset Idiot Colleague Played Taylor Swift to Trigger YouTube Filter." *Vice*, August 10, 2021. www.vice.com/en/article/4avp4m/alameda-county-police-department-taylor-swift-filming-youtube.

Constine, Josh. 2018. "Facebook Will Change Algorithm to Demote 'Borderline Content' That Almost Violates Policies." *TechCrunch*, November 15, 2018. http://social.techcrunch.com/2018/11/15/facebook-borderline-content/.

2019. "Facebook Tries Hiding Like Counts to Fight Envy." *TechCrunch*, September 26, 2019. http://social.techcrunch.com/2019/09/26/facebook-hides-likes/.

Cosens, Barbara, J. B. Ruhl, Niko Soininen, Lance Gunderson, Antti Belinskij, Thorsten Blenckner, Alejandro E. Camacho, et al. 2021. "Governing Complexity: Integrating Science, Governance, and Law to Manage Accelerating Change in the Globalized Commons." *Proceedings of the National Academy of Sciences* 118 (36): e2102798118. https://doi.org/10.1073/pnas.2102798118.

Cox, Joseph. 2019. "Company That Provides Tech to ICE Not Renewing Contract after Code-Deletion Protest." *Vice*, September 23, 2019. www.vice.com/en/article/qvg3q5/chef-not-renewing-ice-immigration-customs-enforcement-contract-after-code-deleting-protest.

Cox, Joseph, and Jason Koebler. 2021. "'FIND THIS FUCK:' Inside Citizen's Dangerous Effort to Cash in on Vigilantism." *Vice*, May 27, 2021. www.vice.com/en/article/y3dpyw/inside-crime-app-citizen-vigilante.

Craig, Robin Kundis, and J. B. Ruhl. 2020. "Adaptive Management for Ecosystem Services across the Wildland-Urban Interface." *International Journal of the Commons* 14 (1): 611–26. https://doi.org/10.5334/ijc.986.

Cusumano, Michael A., Annabelle Gawer, and David B. Yoffie. 2019. *The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power*. New York: Harper Business.

Dagan, Hanoch, and Michael Heller. 2020. "Choice Theory: A Restatement." In *Research Handbook on Private Law Theory*, by Hanoch Dagan and Benjamin C. Zipursky, 112–32. Cheltenham, UK: Edward Elgar.

Dahan, Frederique, Jacek Kubas, Lewis Cohen, and Yana Mihaleva. 2012. "The EBRD's Legal and Regulatory Assessment – What Limits Development of Local Capital Markets Analysis." *Law and Financial Markets Review* 6 (3): 218–23.

De Lara, Yadira Gonzalez, Avner Greif, and Saumitra Jha. 2008. "The Administrative Foundations of Self-Enforcing Constitutions." *American Economic Review* 98 (2): 105–9.

DeNardis, Laura. 2009. *Protocol Politics: The Globalization of Internet Governance. Information Revolution and Global Politics*. Cambridge, MA: MIT Press.

2014. *The Global War for Internet Governance*. New Haven: Yale University Press.

Desrochers, Pierre, and Gert-Jan Hospers. 2007. "Cities and the Economic Development of Nations: An Essay on Jane Jacobs' Contribution to Economic Theory." *Canadian Journal of Regional Science* 30 (1): 115–30.

Dewey, John. 1927. *The Public and Its Problems*. New York: Henry Holt and Company.

DeWitt, Adrienne A. 2010. "The Cost of Free Access to Information: The Controversy over PACER and Open Access to Court Documents." *DttP: Documents to the People* 38 (4): 31–37.

Dhir, Amandeep, Yossiri Yossatorn, Puneet Kaur, and Sufen Chen. 2018. "Online Social Media Fatigue and Psychological Wellbeing—A Study of Compulsive Use, Fear of Missing Out, Fatigue, Anxiety and Depression." *International Journal of Information Management* 40 (June): 141–52. https://doi.org/10.1016/j.ijinfomgt.2018.01.012.

Di Porto, Fabiana, and Marialuisa Zuppetta. 2021. "Co-Regulating Algorithmic Disclosure for Digital Platforms." *Policy and Society* 40 (2): 272–93. https://doi.org/10.1080/14494035.2020.1809052.

Dias Oliva, Thiago, Dennys Marcelo Antonialli, and Alessandra Gomes. 2021. "Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online." *Sexuality & Culture* 25 (2): 700–732. https://doi.org/10.1007/s12119-020-09790-w.

Docquir, Pierre François. 2019. *The Social Media Council: Bringing Human Rights Standards to Content Moderation on Social Media*. Waterloo, ON: Centre for International Governance Innovation. www.cigionline.org/articles/social-media-council-bringing-human-rights-standards-content-moderation-social-media/.

Dolgoplov, Stainslav. 2008. "Insider Trading, Chinese Walls, and Brokerage Commissions: The Origins of Modern Regulation of Information Flows in Securities Markets Insider Trading Symposium – January 27, 2007." *Journal of Law, Economics & Policy* 4 (2): 311–68.

Domino, Jenny. 2020. "Crime as Cognitive Constraint: Facebook's Role in Myanmar's Incitement Landscape and the Promise of International Tort Liability." *Case Western Reserve Journal of International Law* 52 (1): 143–97.

Donahue, John D., and Richard Zeckhauser. 2011. *Collaborative Governance: Private Roles for Public Goals in Turbulent Times*. Princeton, NJ: Princeton University Press.

Doremus, Holly. 2011. "Adaptive Management as an Information Problem." *North Carolina Law Review* 89 (5): 1455–98.

Douek, Evelyn. 2019. "Facebook's 'Oversight Board:' Move Fast with Stable Infrastructure and Humility." *North Carolina Journal of Law and Technology* 21 (1): 1–77.

2021. "The Limits of International Law in Content Moderation." *UC Irvine Journal of International, Transnational, and Comparative Law* 6: 37–76.

Dowding, Keith. 2005. "Is It Rational to Vote? Five Types of Answer and a Suggestion." *The British Journal of Politics and International Relations* 7 (3): 442–59. https://doi.org/10.1111/j.1467-856X.2005.00188.x.

Durkee, Alison. 2019. "Facebook Investors Begin to Revolt against Mark Zuckerberg." *Vanity Fair, June 4, 2019*. www.vanityfair.com/news/2019/06/facebook-investors-revolt-against-mark-zuckerberg.

Dwoskin, Elizabeth, Craig Timberg, and Tony Romm. 2020. "Zuckerberg Once Wanted to Sanction Trump. Then Facebook Wrote Rules That Accommodated Him." *Washington Post*, June 28, 2020. www.washingtonpost.com/technology/2020/06/28/facebook-zuckerberg-trump-hate/.

Edwards, Brian, Elizabeth Naismith Picciani, Sarah Spicer, and Keith Cousins. 2021. "Addressing Rape in Four Minutes or Less: Dating App Reps Left Unprepared to

Respond to Assault Victims." *ProPublica*, May 17, 2021. www.propublica.org/article/addressing-rape-in-four-minutes-or-less-dating-app-reps-left-unprepared-to-respond-to-assault-victims?token=x6pxspak_bzvysutf-lu365b66_2ncxt.

Eichensehr, Kristen E. 2019. "Digital Switzerlands." *University of Pennsylvania Law Review* 167 (3): 665–732.

Elkin-Koren, Niva. 2022. "Government–Platform Synergy and Its Perils." In *Constitutionalising Social Media*, edited by Edoardo Celeste, Amélie Heldt, and Clara Iglesias Keller, 177–98. New York: Bloomsbury.

Ellickson, Robert C. 1991. *Order without Law: How Neighbors Settle Disputes*. 1. Harvard Univ. Press paperback ed.; 3. print, digitally reprinted. Cambridge, MA: Harvard University Press.

Elster, Jon. 2000. *Ulysses Unbound. Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press.

Engelmann, Severin, Jens Grossklags, and Orestis Papakyriakopoulos. 2018. "A Democracy Called Facebook? Participation as a Privacy Strategy on Social Media." In *Privacy Technologies and Policy*, edited by Manel Medina, Andreas Mitrakas, Kai Rannenberg, Erich Schweighofer, and Nikolaos Tsouroulas, 11079:91–108. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02547-2_6.

Engels, Barbara. 2016. "Data Portability Among Online Platforms." *Internet Policy Review* 5 (2): 1–17.

Etter, Lauren. 2017. "What Happens When the Government Uses Facebook as a Weapon?" *Bloomberg Businessweek*, *December 7, 2017*. www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook.

Evans, David S., and Richard Schmalensee. 2010. "Failure to Launch: Critical Mass in Platform Businesses." *Review of Network Economics* 9 (4): 1–26. https://doi.org/10.2202/1446-9022.1256.

Facebook. 2019. "Mark Zuckerberg Stands for Voice and Free Expression." *Meta Newsroom (blog)*. October 17, 2019. https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression/.

Fan, Jenny, and Amy X. Zhang. 2020. "Digital Juries: A Civics-Oriented Approach to Platform Governance." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. Honolulu HI USA: ACM. https://doi.org/10.1145/3313831.3376293.

Fanon, Franz. 1994. *A Dying Colonialism*. Translated by Chevalier, Haakon. New York: Grove Press.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92. https://doi.org/10.2307/2944796.

Ferejohn, John. 1999. "Independent Judges, Dependent Judiciary: Explaining Judicial Independence." *Southern California Law Review* 72 (Issues 2 & 3): 353–84.

Ferreras, Isabelle, and Miranda Richmond Mouillot. 2017. *Firms as Political Entities: Saving Democracy through Economic Bicameralism*. Cambridge, GB; New York, NY: Cambridge University Press.

Fingas, Jon. 2022. "Twitter Says It Inadvertently Ran Ads on Profiles Containing CSAM." *Engadget*, *September 28, 2022*. www.engadget.com/twitter-ads-csam-incident-204104639.html.

Fink, Christina. 2018. "Dangerous Speech, Anti-Muslim Violence, and Facebook in Myanmar." *Journal of International Affairs* 71 (1.5): 43–52.

Fishkin, James S., Robert C. Luskin, and Roger Jowell. 2000. "Deliberative Polling and Public Consultation." *Parliamentary Affairs* 53 (4): 657–66. https://doi.org/10.1093/pa/53.4.657.

Flew, Terry, and Rosalie Gillett. 2021. "Platform Policy: Evaluating Different Responses to the Challenges of Platform Power." *Journal of Digital Media & Policy* 12 (2): 231–46. https://doi.org/10.1386/jdmp_00061_1.

Follesdal, Andreas, and Simon Hix. 2006. "Why There Is a Democratic Deficit in the EU: A Response to Majone and Moravcsik." *JCMS: Journal of Common Market Studies* 44 (3): 533–62. https://doi.org/10.1111/j.1468-5965.2006.00650.x.

Forestal, Jennifer. 2017. "The Architecture of Political Spaces: Trolls, Digital Media, and Deweyan Democracy." *American Political Science Review* 111 (1): 149–61. https://doi.org/10.1017/S0003055416000666.

2021a. "Beyond Gatekeeping: Propaganda, Democracy, and the Organization of Digital Publics." *The Journal of Politics* 83 (1): 306–20. https://doi.org/10.1086/709300.

2021b. "Constructing Digital Democracies: Facebook, Arendt, and the Politics of Design." *Political Studies* 69 (1): 26–44. https://doi.org/10.1177/0032321719890807.

Forte, Andrea, Vanesa Larco, and Amy Bruckman. 2009. "Decentralization in Wikipedia Governance." *Journal of Management Information Systems* 26 (1): 49–72. https://doi.org/10.2753/MIS0742-1222260103.

Freeman, R. Edward, Kirsten Martin, and Bidhan Parmar. 2007. "Stakeholder Capitalism." *Journal of Business Ethics* 74 (4): 303–14. https://doi.org/10.1007/s10551-007-9517-y.

Frenkel, Sheera. 2018. "Microsoft Employees Protest Work with ICE, as Tech Industry Mobilizes over Immigration." *The New York Times*, June 19, 2018, sec. Technology. www.nytimes.com/2018/06/19/technology/tech-companies-immigration-border.html.

Frenkel, Sheera, and Davey Alba. 2021. "In India, Facebook Grapples with an Amplified Version of Its Problems." *The New York Times*, October 23, 2021, sec. Technology. www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html.

Frey, Seth, P. M. Krafft, and Brian C. Keegan. 2019. "'This Place Does What It Was Built For': Designing Digital Institutions for Participatory Change." *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW): 32:1–31. https://doi.org/10.1145/3359134.

Fried, Charles. 2015. *Contract as Promise: A Theory of Contractual Obligation*. Second edition. New York, NY: Oxford University Press.

Frier, Sarah, and Kurt Wagner. 2020. "Facebook Needs Trump Even More Than Trump Needs Facebook." *Bloomberg Businessweek*, September 17, 2020.www.bloomberg.com/news/features/2020-09-17/facebook-and-mark-zuckerberg-need-trump-even-more-than-trump-needs-facebook.

Frischmann, Brett M., Michael J. Madison, and Katherine Jo Strandburg, eds. 2014. *Governing Knowledge Commons*. Oxford; New York: Oxford University Press.

Fuller, Lon L. 1969. "Human Interaction and the Law." *American Journal of Jurisprudence* 14: 1–36.

1978. "The Forms and Limits of Adjudication." Edited by Kenneth I. Winston. *Harvard Law Review* 92 (2): 353–409. https://doi.org/10.2307/1340368.

Fung, Archon, and Erik Olin Wright. 2003. *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance. The Real Utopias Project*, volume 4. London: Verso.

Gadinis, Stavros, and Amelia Miazad. 2021. "A Test of Stakeholder Capitalism." *Journal of Corporation Law* 47 (1): 47–104.

Garcia, David. 2017. "Leaking Privacy and Shadow Profiles in Online Social Networks." *Science Advances* 3 (8): e1701172. https://doi.org/10.1126/sciadv.1701172.

Garcia, David, Pavlin Mavrodiev, and Frank Schweitzer. 2013. "Social Resilience in Online Communities: The Autopsy of Friendster." *Proceedings of the First ACM Conference on Online Social Networks – COSN '13*, 39–50. https://doi.org/10.1145/2512938.2512946.

Gates, Robert M. 2010. "Helping Others Defend Themselves: The Future of U.S. Security Assistance." *Foreign Affairs* 89 (3): 2–6.

Gelter, Martin, and Aurelio Gurrea-Martinez. 2020. "Addressing the Auditor Independence Puzzle: Regulatory Models and Proposal for Reform." *Vanderbilt Journal of Transnational Law* 53 (3): 787–828.

Gergen, Mark P. 1999. "The Jury's Role In Deciding Normative Issues in the American Common Law." *Fordham Law Review* 68 (2): 407–85. https://doi.org/10.2139/ssrn.193748.

Gilbert, David. 2022. "Inside the Far Right QAnon Plot to Overthrow the German Government." *Vice*, December 9, 2022. www.vice.com/en/article/qjk5k7/germany-far-right-qanon-coup-plot.

Gilles, Stephen G. 2001. "On Determining Negligence: Hand Formula Balancing, the Reasonable Person Standard, and the Jury Symposium: The John W. Wade Conference on the Third Restatement of Torts." *Vanderbilt Law Review* 54 (3): 813–62.

Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347–64. https://doi.org/10.1177/1461444809342738.

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

2018b. "Platforms Are Not Intermediaries." *Georgetown Technology Law Review* 2 (2): 198–216.

2018c. "Regulation of and by Platforms." *In The Sage Handbook of Social Media, by Jean Burgess, Alice Marwick, and Thomas Poell*, 30. London: SAGE Publications Ltd.

2022. "Do Not Recommend? Reduction as a Form of Content Moderation." *Social Media + Society* 8 (3): 205630512211175. https://doi.org/10.1177/20563051221117552.

Giuliani, Elisa, and Chiara Macchi. 2014. "Multinational Corporations' Economic and Human Rights Impacts on Developing Countries: A Review and Research Agenda." *Cambridge Journal of Economics* 38 (2): 479–517. https://doi.org/10.1093/cje/bet060.

Gleicher, Nathaniel. 2018. "Coordinated Inauthentic Behavior Explained." *Meta Newsroom* (blog), December 6, 2018. https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/.

Goldman, Eric. 2021. "Content Moderation Remedies." *Michigan Technology Law Review* 28 (1): 1–59.

Goldman, Eric, and Jess Miers. 2021. "Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules." *Journal of Free Speech Law* 1 (Symposium Issue): 191–226.

Goodin, Dan. 2018. "Widely Used Open Source Software Contained Bitcoin-Stealing Backdoor." *Ars Technica*, November 26, 2018. https://arstechnica.com/information-technology/2018/11/hacker-backdoors-widely-used-open-source-software-to-steal-bitcoin/.

Goodman, Ellen P, and Karen Kornbluh. 2021. "The Stakes of User Interface Design for Democracy." *The German Marshall Fund of the United States Policy Paper*. www.gmfus.org/publications/stakes-user-interface-design-democracy.

Gorwa, Robert. 2019a. "What Is Platform Governance?" *Information, Communication & Society* 22 (6): 854–71. https://doi.org/10.1080/1369118X.2019.1573914.

2019b. "The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content." *Internet Policy Review* 8 (2): 1–22. https://doi.org/10.14763/2019.2.1407.

Gorwa, Robert, Reuben Binns, and Christian Katzenbach. 2020. "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance." *Big Data & Society* 7 (1): 2053951719897945. https://doi.org/10.1177/2053951719897945.

Gotanda, Neil. 1991. "A Critique of Our Constitution Is Color-Blind." *Stanford Law Review* 44 (1): 1–68.

Gowder, Paul. 2014a. "Market Unfreedom." *Critical Review* 26 (3–4): 306–47.

2014b. "The Countermajoritarian Complaint." *Transnational Law and Contemporary Problems* 23: 7–33.
2015. "Critical Race Science and Critical Race Philosophy of Science." *Fordham Law Review* 83 (6): 3155–77.
2016. *The Rule of Law in the Real World*. New York: Cambridge University Press.
2018a. "Resisting the Rule of Men." *Saint Louis University Law Journal* 62 (2): 333–59.
2018b. "Transformative Legal Technology and the Rule of Law." *University of Toronto Law Journal* 68 (S1): 82–105.
2019. "Reconstituting We the People: Frederick Douglass and Jürgen Habermas in Conversation." *Northwestern University Law Review* 114 (2): 335–413.
2021. *The Rule of Law in the United States: An Unfinished Project of Black Liberation*. Oxford; New York: Hart.
Gowen, Annie, and Max Bearak. 2017. "Fake News on Facebook Fans the Flames of Hate against the Rohingya in Burma." *Washington Post*, 2017-12-08T06:00-500, sec. Asia & Pacific. www.washingtonpost.com/world/asia_pacific/fake-news-on-facebook-fans-the-flames-of-hate-against-the-rohingya-in-burma/2017/12/07/2c1fe830-ca1f-11e7-b506-8a10ed11ecf5_story.html.
Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–80.
Green, Jessica F. 2014. *Rethinking Private Authority: Agents and Entrepreneurs in Global Environmental Governance*. Princeton: Princeton University Press.
Greenwood, Justin. 2007. "Review Article: Organized Civil Society and Democratic Legitimacy in the European Union." *British Journal of Political Science* 37 (2): 333–57.
Grewal, David Singh. 2008. *Network Power: The Social Dynamics of Globalization*. New Haven: Yale University Press.
Guadagno, Rosanna E., Daniel M. Rempala, Shannon Murphy, and Bradley M. Okdie. 2013. "What Makes a Video Go Viral? An Analysis of Emotional Contagion and Internet Memes." *Computers in Human Behavior* 29 (6): 2312–19. https://doi.org/10.1016/j.chb.2013.04.016.
Gurley, Lauren Kaori. 2019. "Amazon Strikers Say Bezos' Climate Pledge Is 'Not Enough'." *Vice*, September 19, 2019. www.vice.com/en/article/wjwgg5/amazon-strikers-say-bezos-climate-pledge-is-not-enough.
Guynn, Jessica. 2021. "'Burn down DC': Violence That Erupted at Capitol Was Incited by Pro-Trump Mob on Social Media." *USA Today*, January 6, 2021. www.usatoday.com/story/tech/2021/01/06/trump-riot-twitter-parler-proud-boys-boogaloos-antifa-qanon/6570794002/.
Habermas, Jürgen. 2001. "Constitutional Democracy: A Paradoxical Union of Contradictory Principles?" *Political Theory* 29 (6): 766–81. https://doi.org/10.1177/0090591701029006002.
Hadfield, Gillian, and Barry Weingast. 2013. "Law without the State: Legal Attributes and the Coordination of Decentralized Collective Punishment." *Journal of Law & Courts* 1 (1): 3–34.
2014. "Microfoundations of the Rule of Law." *Annual Review of Political Science* 17: 21–42.
Hagey, Keach, and Jeff Horwitz. 2021. "Facebook's Internal Chat Boards Show Politics Often at Center of Decision Making." *Wall Street Journal*, October 24, 2021, sec. Tech. www.wsj.com/articles/facebook-politics-decision-making-documents-11635100195.
Haggard, Stephan, Andrew MacIntyre, and Lydia Tiede. 2008. "The Rule of Law and Economic Development." *Annual Review of Political Science* 11: 205–34.
Haggart, Blayne. 2020. "Global Platform Governance and the Internet-Governance Impossibility Theorem." *Journal of Digital Media & Policy* 11 (3): 321–39. https://doi.org/10.1386/jdmp_00028_1.

Haggart, Blayne, and Clara Iglesias Keller. 2021. "Democratic Legitimacy in Global Platform Governance." *Telecommunications Policy* 45 (6): 102152. https://doi.org/10.1016/j.telpol.2021.102152.

Haggart, Blayne, Natasha Tusikov, and Jan Aart Scholte, eds. 2021. Power and Authority in Internet Governance: Return of the State? Routledge Global Cooperation Series. London; New York, NY: Routledge.

Haimson, Oliver L., and Anna Lauren Hoffmann. 2016. "Constructing and Enforcing 'Authentic' Identity Online: Facebook, Real Names, and Non-Normative Identities." *First Monday*, June. https://doi.org/10.5210/fm.v21i6.6791.

Halliday, Josh. 2012. "Twitter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party'." *The Guardian*, March 22, 2012. www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech.

Hao, Karen. 2021. "How Facebook Got Addicted to Spreading Misinformation." *MIT Technology Review*, March 11, 2021. www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/.

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162 (3859): 1243–48.

Harnett, Sam. 2018. "Cities Made Millions Selling Taxi Medallions, Now Drivers Are Paying the Price." National Public Radio. October 15, 2018. www.npr.org/2018/10/15/656595597/cities-made-millions-selling-taxi-medallions-now-drivers-are-paying-the-price.

Harwell, Drew. 2021. "As GameStop Stock Crumbles, Newbie Traders Reckon with Heavy Losses." *Washington Post*, February 2, 2021. www.washingtonpost.com/technology/2021/02/02/gamestop-stock-plunge-losers/.

2022. "A Fake Tweet Sparked Panic at Eli Lilly and May Have Cost Twitter Millions." *Washington Post*, November 14, 2022. www.washingtonpost.com/technology/2022/11/14/twitter-fake-eli-lilly/.

Harwell, Drew, and Nitasha Tiku. 2020. "Google's Star AI Ethics Researcher, One of a Few Black Women in the Field, Says She Was Fired for a Critical Email." *Washington Post*, December 3, 2020. www.washingtonpost.com/technology/2020/12/03/timnit-gebru-google-fired/.

Hatmaker, Taylor. 2018. "Zuckerberg Denies Knowledge of Facebook Shadow Profiles." *TechCrunch* (blog). April 11, 2018. https://techcrunch.com/2018/04/11/facebook-shadow-profiles-hearing-lujan-zuckerberg/.

2021. "Facebook Oversight Board Says Other Social Networks 'Welcome to Join' if Project Succeeds." *TechCrunch* (blog). February 11, 2021. https://social.techcrunch.com/2021/02/11/facebook-oversight-board-other-social-networks-beyond-facebook/.

Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.

Haynes, Brad. 2016. "Facebook Executive Jailed in Brazil as Court Seeks WhatsApp Data." *Reuters*, March 1, 2016, sec. Media and Telecoms. www.reuters.com/article/us-facebook-brazil-idUSKCN0W34WF.

Hearn, Alex. 2017. "Google Acts against Fake News on Search Engine." *The Guardian*, April 25, 2017, sec. Technology. www.theguardian.com/technology/2017/apr/25/google-launches-major-offensive-against-fake-news.

Helberger, Natali, Jo Pierson, and Thomas Poell. 2018. "Governing Online Platforms: From Contested to Cooperative Responsibility." *The Information Society* 34 (1): 1–14. https://doi.org/10.1080/01972243.2017.1391913.

Hern, Alex. 2020. "WhatsApp to Impose New Limit on Forwarding to Fight Fake News." *The Guardian*, April 7, 2020, sec. Technology. www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news.

Herrman, John, and Mike Isaac. 2016. "Conservatives Accuse Facebook of Political Bias." *The New York Times*, May 10, 2016, sec. Technology. www.nytimes.com/2016/05/10/technology/conservatives-accuse-facebook-of-political-bias.html.

Hibbing, John R., and Elizabeth Theiss-Morse. 2008. "Voice, Validation, and Legitimacy." In *Cooperation: The Political Psychology of Effective Human Interaction*, edited by Brandon A. Sullivan, Mark Snyder, and John L. Sullivan, 123–42. Malden, MA: Blackwell Pub.

Hiller, Janine S., and Scott J. Shackelford. 2018. "The Firm and Common Pool Resource Theory: Understanding the Rise of Benefit Corporations." *American Business Law Journal* 55 (1): 5–51. https://doi.org/10.1111/ablj.12116.

Hofmann, Jeanette. 2016. "Multi-Stakeholderism in Internet Governance: Putting a Fiction into Practice." *Journal of Cyber Policy* 1 (1): 29–49. https://doi.org/10.1080/23738871.2016.1158303.

Hogan, Libby, and Michael Safi. 2018. "Revealed: Facebook Hate Speech Exploded in Myanmar during Rohingya Crisis." *The Guardian*, April 3, 2018, sec. World News. www.theguardian.com/world/2018/apr/03/revealed-facebook-hate-speech-exploded-in-myanmar-during-rohingya-crisis.

Hollister, Sean. 2021. "FCC Chairman Ajit Pai Gave up on His Legally Dicey Attempt to 'Clarify' Internet Law." The Verge. January 7, 2021. www.theverge.com/2021/1/7/22219677/fcc-ajit-pai-section-230-its-over.

Holpuch, Amanda. 2015. "Facebook Adjusts Controversial 'Real Name' Policy in Wake of Criticism." *The Guardian*, December 15, 2015, sec. Technology. www.theguardian.com/us-news/2015/dec/15/facebook-change-controversial-real-name-policy.

Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385–89. https://doi.org/10.1073/pnas.0403723101.

Horwitz, Jeff. 2021. "Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That's Exempt." *Wall Street Journal*, September 13, 2021, sec. Tech. www.wsj.com/articles/facebook-files-xcheck-zuckerberg-elite-rules-11631541353.

Horwitz, Jeff, and Newley Purnell. 2021. "India Threatens Jail for Facebook, WhatsApp and Twitter Employees." *Wall Street Journal*, March 5, 2021, sec. Tech. www.wsj.com/articles/india-threatens-jail-for-facebook-whatsapp-and-twitter-employees-11614964542.

Horwitz, Jeff, and Justin Scheck. 2021. "Facebook Increasingly Suppresses Political Movements It Deems Dangerous." *Wall Street Journal*, October 22, 2021, sec. Tech. www.wsj.com/articles/facebook-suppresses-political-movements-patriot-party-11634937358.

Horwitz, Jeff, and Deepa Seetharaman. 2020. "Facebook Executives Shut Down Efforts to Make the Site Less Divisive." *Wall Street Journal*, May 26, 2020, sec. Tech. www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.

Hoseini, Mohamad, Philipe Melo, Fabricio Benevenuto, Anja Feldmann, and Savvas Zannettou. 2021. "On the Globalization of the QAnon Conspiracy Theory Through Telegram." *arXiv*. https://doi.org/10.48550/arXiv.2105.13020.

Hovenkamp, Herbert. 2021. "Antitrust and Platform Monopoly." *Yale Law Journal* 130 (8): 1952–2050.

Huffman, Stephen. 2000. "The Navajo Code Talkers: A Cryptologic and Linguistic Perspective." *Cryptologia* 24 (4): 289–320. https://doi.org/10.1080/01611190008984247.

Hunt, Melissa G., Rachel Marx, Courtney Lipson, and Jordyn Young. 2018. "No More FOMO: Limiting Social Media Decreases Loneliness and Depression." *Journal of Social and Clinical Psychology* 37 (10): 751–68. https://doi.org/10.1521/jscp.2018.37.10.751.

Huth, Paul K. 1999. "Deterrence and International Conflict: Empirical Findings and Theoretical Debates." *Annual Review of Political Science* 2 (1): 25–48. https://doi.org/10.1146/annurev.polisci.2.1.25.

India Times. 2020. "Meet Vijaya Gadde, an Indian-Born Twitter Head Who Decides on Blocking Tweets, Users," January 16, 2020.

Ingram, Mathew. 2019. "The Facebook Supreme Court Will See You Now." *Columbia Journalism Review*, September 19, 2019. www.cjr.org/the_media_today/facebook-supreme-court.php.

Isaac, Mike, and Kate Conger. 2021. "Reddit Bans Forum Dedicated to Supporting Trump, and Twitter Permanently Suspends His Allies Who Spread Conspiracy Theories." *The New York Times*, January 8, 2021, sec. U.S. www.nytimes.com/2021/01/08/us/politics/reddit-bans-forum-dedicated-to-supporting-trump-and-twitter-permanently-suspends-his-allies-who-spread-conspiracy-theories.html.

Isaac, Mike, Cecilia Kang, and Sheera Frenkel. 2020. "Zuckerberg Defends Hands-Off Approach to Trump's Posts." *The New York Times*, June 2, 2020, sec. Technology. www.nytimes.com/2020/06/02/technology/zuckerberg-defends-facebook-trump-posts.html.

Isaac, T.M. Thomas, and Patrick Heller. 2003. "Democracy and Development: Decentralized Planning in Kerala." In *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*, edited by Archon Fung and Erik Olin Wright, 77–110. The real utopias project, volume 4. London: Verso.

Jacobs, Jane. 1992. *The Death and Life of Great American Cities*. New York: Vintage Books.

2016. *Vital Little Plans: The Short Works of Jane Jacobs*. Edited by Samuel Zipp and Nathan Storring. First edition. New York: Random House.

Jin, Dal Yong. 2013. "The Construction of Platform Imperialism in the Globalization Era." *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 11 (1): 145–72. https://doi.org/10.31269/triplec.v11i1.458.

2017. *Digital Platforms, Imperialism and Political Culture*.

Kadens, Emily. 2012. "The Myth of the Customary Law Merchant." *Texas Law Review* 90 (5): 1153–1206.

2015. "The Medieval Law Merchant: The Tyranny of a Construct." *Journal of Legal Analysis* 7 (2): 251–89. https://doi.org/10.1093/jla/lav004.

Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. 1st pbk. ed. New York: Farrar, Straus and Giroux.

Kaminski, Margot E. 2014. "The Capture of International Intellectual Property Law through the U.S. Trade Regime." *Southern California Law Review* 87 (4): 977–1052.

Kantrowitz, Alex. 2019. "Facebook Is Still Prioritizing Scale over Safety." *BuzzFeed News*, December 17, 2019. www.buzzfeednews.com/article/alexkantrowitz/after-years-of-scandal-facebooks-unhealthy-obsession-with.

Kaye, David. 2019a. "Legitimacy, Collective Authority and Internet Governance: A Reflection on David Caron's Study of the UN Security Council." *Ecology Law Quarterly* 46 (1): 135–48. https://doi.org/10.15779/Z38M902363.

2019b. *Speech Police: The Global Struggle to Govern the Internet*. New York: Columbia Global Reports.

Keller, Clara Iglesias. 2022. "The Perks of Co-Regulation: An Institutional Arrangement for Social Media Regulation?" In *Constitutionalising social media*, edited by Edoardo Celeste, Amélie Heldt, and Clara Iglesias Keller, 217–34.

Keller, Daphne. 2021. "The Future of Platform Power: Making Middleware Work." *Journal of Democracy* 32 (3): 168–72. https://doi.org/10.1353/jod.2021.0043.

Kemp, Catherine. 1999. "Habermas among the Americans: Some Reflections on the Common Law Fifth Annual Legal Theory Symposium: Exploring Habermas on Law and Democracy." *Denver University Law Review* 76 (4): 961–76.

Kettemann, Matthias C., and Fertmann, Martin. 2021. "Platform-Proofing Democracy: Social Media Councils as Tools to Increase the Public Accountability of Online Platforms." *Friedrich Neumann Foundation for Freedom*. https://shop.freiheit.org/download/P2@1084/444384/2021_FNF_Platformräte_Social%20Media%20Councils_eng%20final.pdf.

Kiesler, Sara, Robert E. Kraut, Paul Resnick, and Aniket Kittur. 2011. "Regulating Behavior in Online Communities." In *Building Successful Online Communities: Evidence-Based Social Design*, edited by Robert E. Kraut and Paul Resnick, 125–77. Cambridge, Mass: MIT Press.

Kilgour, Marc D., and Frank C. Zagare. 1991. "Credibility, Uncertainty, and Deterrence." *American Journal of Political Science* 35 (2): 305–34. https://doi.org/10.2307/2111365.

Kim, Eugene. 2019. "Amazon Added a First-Ever Warning about Counterfeit Products to Its Earnings Report." *CNBC*. February 4, 2019. www.cnbc.com/2019/02/04/amazon-10k-warns-investors-about-counterfeit-problem-for-first-time.html.

Klein, Scott, and Jeff Kao. 2021. "Why We Published More Than 500 Videos Taken by Parler Users of the Capitol Riot." *ProPublica*, January 17, 2021. www.propublica.org/article/why-we-published-parler-users-videos-capitol-attack?token=Tu5C70R2pCBv8Yj33AkMh2E-mHz3d6iu.

Klijn, Erik Hans, and Jurian Edelenbos. 2013. "The Influence of Democratic Legitimacy on Outcomes in Governance Networks." *Administration & Society* 45 (6): 627–50. https://doi.org/10.1177/0095399712454113.

Klonick, Kate. 2018. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131 (6): 1598–1670.

2019. "Does Facebook's Oversight Board Finally Solve the Problem of Online Speech?" Centre for International Governance Innovation. www.cigionline.org/articles/does-facebooks-oversight-board-finally-solve-problem-online-speech.

2020. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression." *Yale Law Journal* 129 (8): 2418–99.

Kohler-Koch, Beate. 2010. "Civil Society and EU Democracy: 'Astroturf' Representation?" *Journal of European Public Policy* 17 (1): 100–116. https://doi.org/10.1080/13501760903464986.

Kolhatkar, Sheelah. 2019. "How Elizabeth Warren Came Up with a Plan to Break Up Big Tech." *The New Yorker*, August 20, 2019. www.newyorker.com/business/currency/how-elizabeth-warren-came-up-with-a-plan-to-break-up-big-tech.

Koontz, Tomas M., Divya Gupta, Pranietha Mudliar, and Pranay Ranjan. 2015. "Adaptive Institutions in Social-Ecological Systems Governance: A Synthesis Framework." *Environmental Science & Policy* 53 (November): 139–51. https://doi.org/10.1016/j.envsci.2015.01.003.

Kosseff, Jeff. 2019. "First Amendment Protection for Online Platforms." *Computer Law & Security Review* 35 (2): 199–213. https://doi.org/10.1016/j.clsr.2018.12.002.

Kraut, Robert E., Paul Resnick, and Sara Kiesler. 2011. *Building Successful Online Communities: Evidence-Based Social Design*. Cambridge, MA: MIT Press.

Kulp, Patrick. 2019. "Disney, Nestle and Fortnite Publisher Pull YouTube Ads Following Child Predator Controversy." *Adweek*, February 20, 2019. www.adweek.com/programmatic/fortnite-publisher-pulls-pre-roll-ads-on-youtube-following-child-exploitation-controversy/.

Kuran, Timur. 1989. "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution." *Public Choice* 61 (1): 41–74.
1991. "The East European Revolution of 1989: Is It Surprising That We Were Surprised?" *The American Economic Review* 81 (2): 121–25.
Kwet, Michael. 2019. "Digital Colonialism: US Empire and the New Imperialism in the Global South." *Race & Class* 60 (4): 3–26.
Kymlicka, Will. 2015. "Solidarity in Diverse Societies: Beyond Neoliberal Multiculturalism and Welfare Chauvinism." *Comparative Migration Studies* 3 (1): 17. https://doi.org/10.1186/s40878-015-0017-4.
Land, Molly K. 2013. "Toward an International Law of the Internet." *Harvard International Law Journal* 54 (2): 393–458.
2019. "Against Privatized Censorship: Proposals for Responsible Delegation." *Virginia Journal of International Law* 60 (2): 363–431. https://doi.org/10.2139/ssrn.3442184.
Landemore, Hélène. 2013. "Deliberation, Cognitive Diversity, and Democratic Inclusiveness: An Epistemic Argument for the Random Selection of Representatives." *Synthese* 190 (7): 1209–31. https://doi.org/10.1007/s11229-012-0062-6.
2020. *Open Democracy: Reinventing Popular Rule for the Twenty-First Century*. Princeton: Princeton University Press.
Landemore, Hélène, and Isabelle Ferreras. 2016. "In Defense of Workplace Democracy: Towards a Justification of the Firm–State Analogy." *Political Theory* 44 (1): 53–81. https://doi.org/10.1177/0090591715600035.
Lange, Ann-Christina, Marc Lenglet, and Robert Seyfert. 2016. "Cultures of High-Frequency Trading: Mapping the Landscape of Algorithmic Developments in Contemporary Financial Markets." *Economy and Society* 45 (2): 149–65. https://doi.org/10.1080/03085147.2016.1213986.
Langston, Jennifer. 2018. "How PhotoDNA for Video Is Being Used to Fight Online Child Exploitation." *Microsoft/On the Issues* (blog). September 12, 2018. https://news.microsoft.com/on-the-issues/2018/09/12/how-photodna-for-video-is-being-used-to-fight-online-child-exploitation/.
Langvardt, Kyle. 2021. "Can the First Amendment Scale?" *Journal of Free Speech Law* 1 (Symposium Issue): 273–302.
Law, David S. 2009. "A Theory of Judicial Power and Judicial Review." *Georgetown Law Journal* 97 (3): 723–802.
Leader Maynard, Jonathan, and Susan Benesch. 2016. "Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Prevention." Genocide Studies and Prevention: An International Journal 9 (3): 70–95. http://dx.doi.org/10.5038/1911-9933.9.3.1317
Leary, Alex, and John D. McKinnon. 2020. "Trump Considers Forming Panel to Review Complaints of Online Bias." *Wall Street Journal*, May 23, 2020, sec. Politics. www.wsj.com/articles/trump-considers-forming-panel-to-review-complaints-of-online-bias-11590238800.
Lee, Ronan. 2019. "Extreme Speech in Myanmar: The Role of State Media in the Rohingya Forced Migration Crisis." *International Journal of Communication* 13: 3203–24.
Lessig, Lawrence. 1999. *Code and Other Laws of Cyberspace*. New York: Basic Books.
2003. "Dunwody Distinguished Lecture in Law: The Creative Commons." *Florida Law Review* 55 (3): 763–78.
Levy, Jacob T. 2017. *Rationalism, Pluralism, and Freedom*.
Levy, Steven. 2020a. *Facebook: The Inside Story*. New York: Dutton.
2020b. "Why Mark Zuckerberg's Oversight Board May Kill His Political Ad Policy." *Wired*, January 28, 2020. www.wired.com/story/facebook-oversight-board-bylaws/.

Lima, Lucas, Julio C. S. Reis, Philipe Melo, Fabricio Murai, Leandro Araújo, Pantelis Vikatos, and Fabrício Benevenuto. 2018. "Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System." *arXiv:1807.03688 [Cs]*, July.

Liu, Lizhi, and Barry R. Weingast. 2018. "Taobao, Federalism, and the Emergence of Law, Chinese Style." *Minnesota Law Review* 102 (4): 1563–90.

Liu, Yi, Pinar Yildirim, and Z. John Zhang. 2021. "Social Media, Content Moderation, and Technology." *arXiv:2101.04618 [Econ, q-Fin]*, January.

Lobel, Orly. 2016. "The Law of the Platform." *Minnesota Law Review* 101 (1): 87–166.

Loo, Rory van. 2020. "The New Gatekeepers: Private Firms as Public Enforcers." *Virginia Law Review* 106 (2): 467–522.

Looft, Chris, and Layla Ferris. 2021. "Facebook Whistleblower Documents Offer New Revelations about Jan. 6 Response." *ABC News*, October 24, 2021. https://abcnews.go.com/Technology/facebook-whistleblower-documents-offer-revelations-jan-response/story?id=80694096.

Loos, Marco, and Josia Luzak. 2021. "Update the Unfair Contract Terms Directive for Digital Services." PE 676.006. EPRS Studies. European Parliament. www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2021)676006.

Lundgren, Björn, and H. Orri Stefánsson. 2020. "Against the De Minimis Principle." *Risk Analysis* 40 (5): 908–14. https://doi.org/10.1111/risa.13445.

Lux, Dottie. 2017. "Facebook's Hate Speech Policies Censor Marginalized Users." *Wired*, August 14, 2017. www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/.

Lymperopoulos, Ilias N., and George D. Ioannou. 2016. "Understanding and Modeling the Complex Dynamics of the Online Social Networks: A Scalable Conceptual Approach." *Evolving Systems* 7 (3): 207–32. https://doi.org/10.1007/s12530-016-9145-9.

Lynch, Kevin. 1960. *The Image of the City*. Cambridge, MA: MIT Press.

Lyons, Tessa. 2018. "Hard Questions: How Is Facebook's Fact-Checking Program Working?" Facebook Newsroom. June 14, 2018. https://about.fb.com/news/2018/06/hard-questions-fact-checking/.

Mac, Ryan, and Craig Silverman. 2021. "'Mark Changed the Rules': How Facebook Went Easy on Alex Jones and Other Right-Wing Figures." *BuzzFeed News*, February 21, 2021. www.buzzfeednews.com/article/ryanmac/mark-zuckerberg-joel-kaplan-facebook-alex-jones.

Mac, Ryan, Craig Silverman, and Jane Lytvynenko. 2021. "Facebook Stopped Employees from Reading an Internal Report about Its Role in the Insurrection. You Can Read It Here." *BuzzFeed News*, April 26, 2021. www.buzzfeednews.com/article/ryanmac/full-facebook-stop-the-steal-internal-report.

MacPhail, James R. N. 1926. "The Royal Revenues of England, 1066-1399 (Book Review)." *The Scottish Historical Review* 23 (92): 250–53.

Madrigal, Alexis C. 2017. "What Facebook Did to American Democracy." *The Atlantic*, October 12, 2017. www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/.

Mahtani, Shibani, and Regine Cabato. 2019. "Why Crafty Internet Trolls in the Philippines May Be Coming to a Website Near You." *Washington Post*, July 26, 2019. www.washingtonpost.com/world/asia_pacific/why-crafty-internet-trolls-in-the-philippines-may-be-coming-to-a-website-near-you/2019/07/25/c5d42ee2-5c53-11e9-98d4-844088d135f2_story.html.

Marantz, Andrew. 2020. "Why Facebook Can't Fix Itself." *The New Yorker*, October 19, 2020. www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself.

Marcello, Maria Carolina. 2021. "Brazil Senate Leader Kills Bolsonaro Decree Criticized by Tech Firms." *Reuters*, September 15, 2021, sec. *Americas*. www.reuters.com/world/americas/brazil-senate-president-kills-bolsonaros-social-media-decree-source-says-2021-09-14/.

March, James G., and Johan P. Olsen. 1984. "The New Institutionalism: Organizational Factors in Political Life." *The American Political Science Review* 78 (3): 734–49. https://doi.org/10.2307/1961840.

Marks, Gary. 1996. "An Actor-Centred Approach to Multi-Level Governance." *Regional & Federal Studies* 6 (2): 20–38. https://doi.org/10.1080/13597569608420966.

Marsden, Chris, Trisha Meyer, and Ian Brown. 2020. "Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?" *Computer Law & Security Review* 36 (April): 105373. https://doi.org/10.1016/j.clsr.2019.105373.

Martin-Laborda, Antonio Robles. 2017. "Merger Control and Online Platforms: The Relevance of Network Effects." *Market and Competition Law Review* 1 (2): 69–100.

Marwick, Alice E., and danah boyd. 2011. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13 (1): 114–33. https://doi.org/10.1177/1461444810365313.

Massachi, Sahar. 2021. "How to Save Our Social Media by Treating It Like a City." *MIT Technology Review*, December 20, 2021. www.technologyreview.com/2021/12/20/1042709/how-to-save-social-media-treat-it-like-a-city/.

Mattei, Ugo, and Laura Nader. 2008. *Plunder: When the Rule of Law Is Illegal*. Malden, Mass: Blackwell Pub.

Mattioli, Dana. 2020. "Amazon Scooped Up Data from Its Own Sellers to Launch Competing Products." *Wall Street Journal*, April 24, 2020, sec. Tech. www.wsj.com/articles/amazon-scooped-up-data-from-its-own-sellers-to-launch-competing-products-11587650015.

May, Christopher. 2004. "Capacity Building and the (Re)production of Intellectual Property Rights." *Third World Quarterly* 25 (5): 821–37. https://doi.org/10.1080/01436590420002319 74.

Márkus, György. 1983. "Concepts of Ideology in Marx." *Canadian Journal of Political and Social Theory* 7 (1-2): 84–103.

McClean, Emma. 2008. "The Responsibility to Protect: The Role of International Human Rights Law." *Journal of Conflict and Security Law* 13 (1): 123–52. https://doi.org/10.1093/jcsl/krn012.

McGregor, Shannon C., and Daniel Kreiss. 2019. "Conservatives Say Google and Facebook Are Censoring Them. Here's the Real Background." *Washington Post/Monkey Cage Blog*, August 1, 2019. www.washingtonpost.com/politics/2019/08/01/are-google-facebook-censoring-conservatives-problem-is-more-widespread-than-that/.

McPherson, Poppy. 2020. "Facebook Rejects Request to Release Myanmar Officials' Data for Genocide Case." *Reuters*, August 6, 2020. www.reuters.com/article/us-myanmar-facebook-idUSKCN2521PI.

Meade, Amanda, Josh Taylor, and Daniel Hurst. 2021. "Facebook Reverses Australia News Ban After Government Makes Media Code Amendments." *The Guardian*, February 23, 2021, sec. Media. www.theguardian.com/media/2021/feb/23/facebook-reverses-australia-news-ban-after-government-makes-media-code-amendments.

Meng, Anne. 2020. *Constraining Dictatorship: From Personalized Rule to Institutionalized Regimes*. Political Economy of Institutions and Decisions. Cambridge; New York, NY: Cambridge University Press.

Menkhaus, Ken. 2014. "State Failure, State-Building, and Prospects for a 'Functional Failed State' in Somalia." *The ANNALS of the American Academy of Political and Social Science* 656 (1): 154–72. https://doi.org/10.1177/0002716214547002.

Mezrich, Ben. 2021. *The Antisocial Network: The GameStop Short Squeeze and the Ragtag Group of Amateur Traders That Brought Wall Street to Its Knees*. First edition. New York: Grand Central Publishing.

Mérieau, Eugénie. 2019. "Thailand's Lèse-Majesté Law: On Blasphemy in a Buddhist Kingdom." *Buddhism, Law & Society* 4: 53–92.

2021. "A History of the Thai Lèse-Majesté Law." In *Thai Legal History*, edited by Andrew Harding and Munin Pongsapan, 1st ed., 77–88. Cambridge University Press. https://doi.org/10.1017/9781108914369.007.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics & Politics* 2 (1): 1–23. https://doi.org/10.1111/j.1468-0343.1990.tb00020.x.

Miller, John H., and Scott E. Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press.

Mkandawire, Thandika. 2002. "Incentives, Governance, and Capacity Development in Africa." *African Issues* 30 (1): 15–20. https://doi.org/10.1017/S1548450500006247.

Moe, Terry M. 1984. "The New Economics of Organization." *American Journal of Political Science* 28 (4): 739–77. https://doi.org/10.2307/2110997.

Molitorisz, Sacha, James Meese, and Jennifer Hagedorn. 2021. "From Shadow Profiles to Contact Tracing: Qualitative Research into Consent and Privacy." *Law, Technology and Humans* 3 (2): 46–60.

Moog, Sandra, André Spicer, and Steffen Böhm. 2015. "The Politics of Multi-Stakeholder Initiatives: The Crisis of the Forest Stewardship Council." *Journal of Business Ethics* 128 (3): 469–93. https://doi.org/10.1007/s10551-013-2033-3.

Morell, Mayo Fuster. 2014. "Governance of Online Creation Communities for the Building of Digital Commons: Viewed through the Framework of Institutional Analysis and Development." In *Governing Knowledge Commons*, edited by Brett M. Frischmann, Michael J. Madison, and Katherine Jo Strandburg, 281–311. Oxford; New York: Oxford University Press.

Moroni, Stefano. 2016. "Urban Density after Jane Jacobs: The Crucial Role of Diversity and Emergence." *City, Territory and Architecture* 3 (1): art. 13. https://doi.org/10.1186/s40410-016-0041-1.

Morozov, Evgeny. 2013. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs.

Mozilla Foundation. 2021. "Mozilla Investigation: YouTube Algorithm Recommends Videos That Violate the Platform's Very Own Policies." Mozilla Foundation. July 7, 2021. https://foundation.mozilla.org/en/blog/mozilla-investigation-youtube-algorithm-recommends-videos-that-violate-the-platforms-very-own-policies/.

Mozur, Paul. 2018. "A Genocide Incited on Facebook, with Posts from Myanmar's Military." *The New York Times*, October 15, 2018, sec. Technology. www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Mughan, Siân. 2021. "Municipal Reliance on Fine and Fee Revenues: How Local Courts Contribute to Extractive Revenue Practices in US Cities." *Public Budgeting & Finance* 41 (2): 22–44. https://doi.org/10.1111/pbaf.12277.

Murphy, Hannah. 2019. "Inside Facebook's Information Warfare Team." *Financial Times*, July 6, 2019.

Müller, Jan-Werner. 2007a. *Constitutional Patriotism*. Princeton: Princeton University Press.

2007b. "Is Europe Converging on Constitutional Patriotism? (And if So: Is It Justified?)." *Critical Review of International Social and Political Philosophy* 10 (3): 377–87.

Myers, Steven Lee, and Sheera Frenkel. 2022. "How Russian Propaganda Is Reaching Beyond English Speakers." *The New York Times*, August 9, 2022, sec. Business. www.nytimes.com/2022/08/09/business/russia-propaganda-spanish-social-media.html.

Naīm, 'Abd Allāh Aḥmad. 2021. *Decolonizing Human Rights*. Cambridge; New York, NY: Cambridge University Press.

Napoli, Philip M. 2014. "Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption: Automated Media." *Communication Theory* 24 (3): 340–60. https://doi.org/10.1111/comt.12039.

2015. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." *Telecommunications Policy*, *SPECIAL ISSUE ON THE GOVERNANCE OF SOCIAL MEDIA*, 39 (9): 751–60. https://doi.org/10.1016/j.telpol.2014.12.003.

Newbury, Catharine. 1998. "Ethnicity and the Politics of History in Rwanda." *Africa Today* 45 (1): 7–24.

Newton, Casey. 2019. "The Trauma Floor: The Secret Lives of Facebook Moderators in America." *The Verge*, February 25, 2019. www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.

2022. "How Facebook Undercut the Oversight Board." *The Verge*, *May* 12, 2022. www.theverge.com/23068243/facebook-meta-oversight-board-putin-russia-ukraine-decision.

Nicholson, Helen J. 2021. *The Knights Templar*. Stroud, UK: Sutton.

Nix, Naomi. 2022. "Facebook Bans Hate Speech but Still Makes Money from White Supremacists." *Washington Post*, August 10, 2022. www.washingtonpost.com/technology/2022/08/10/facebook-white-supremacy-ads/.

North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *The Journal of Economic History* 49 (4): 803–32. https://doi.org/10.1017/S0022050700009451.

North, Douglass C., Barry R. Weingast, and John Joseph Wallis. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Nunez, Michael. 2016. "Former Facebook Workers: We Routinely Suppressed Conservative News." *Gizmodo* (blog). May 9, 2016. https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006.

O'Kane, Caitlin. 2020. "Facebook Page That Copies President Trump's Posts Gets Flagged for Violence – When the President's Didn't." CBS News, June 12, 2020. www.cbsnews.com/news/facebook-donald-trump-copy-account-flagged-inciting-violence/.

O'Reilly, Tim. 2011. "Government as a Platform." *Innovations: Technology, Governance, Globalization* 6 (1): 13–40. https://doi.org/10.1162/INOV_a_00056.

Ober, Josiah. 2008. *Democracy and Knowledge: Innovation and Learning in Classical Athens. Innovation and Learning in Classical Athens*. Princeton: Princeton University Press.

Okin, Susan Moller. 1999. *Is Multiculturalism Bad for Women?* Princeton, NJ: Princeton University Press.

Olshan, Jeremy. 2006. "Taxi Tycoon Pays $25m Fare for Medallion Battalion." *New York Post*, June 17, 2006. https://nypost.com/2006/06/17/taxi-tycoon-pays-25m-fare-for-medallion-battalion/.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *The American Political Science Review* 87 (3): 567–76. https://doi.org/10.2307/2938736.

Ostrom, Elinor. 2003. "How Types of Goods and Property Rights Jointly Affect Collective Action." *Journal of Theoretical Politics* 15 (3): 239–70. https://doi.org/10.1177/0951692803015003002.

2009. "A Polycentric Approach for Coping with Climate Change." *Background Paper to the 2010 World Development Report Policy Research Working Paper 5095*. World Bank.
2010a. "Beyond Markets and States: Polycentric Governance of Complex Economic Systems." *American Economic Review* 100 (3): 641–72.
2010b. "Polycentric Systems for Coping with Collective Action and Global Environmental Change." *Global Environmental Change* 20 (4): 550–57. https://doi.org/10.1016/j.gloenvcha.2010.07.004.
2015. *Governing the Commons: The Evolution of Institutions for Collective Action*. *Canto Classics*. Cambridge, United Kingdom: Cambridge University Press.
Pahwa, Nitish. 2021. "The Facebook Crisis in India Might Be the Worst Facebook Crisis of All." *Slate Future Tense*, October 26, 2021. https://slate.com/technology/2021/10/facebook-papers-india-modi-misinformation-rss-bjp.html.
Pasquale, Frank. 2018. "Tech Platforms and the Knowledge Problem." *American Affairs* 2 (2): 3–16.
Passini, Stefano. 2012. "The Facebook and Twitter Revolutions: Active Participation in the Twenty-First Century." *Human Affairs* 22 (3): 301–12. https://doi.org/10.2478/s13374-012-0025-0.
Patten, Alan. 2012. "Liberal Neutrality: A Reinterpretation and Defense*." *Journal of Political Philosophy* 20 (3): 249–72. https://doi.org/10.1111/j.1467-9760.2011.00406.x.
Pearson, Erika, and Adon C. M. Moskal. 2017. "U Tried!: Failure in a University Social Network Site." In *Studying Digital Media Audiences: Perspectives from Australasia*, edited by Ramaswami Harindranath and Craig Hight, 116–32. New York, NY: Routledge.
Percy, Sarah, and Anja Shortland. 2013. "The Business of Piracy in Somalia." *Journal of Strategic Studies* 36 (4): 541–78. https://doi.org/10.1080/01402390.2012.750242.
Perel, Maayan, and Niva Elkin-Koren. 2016. "Accountability in Algorithmic Copyright Enforcement." *Stanford Technology Law Review* 19 (3): 473–532.
Pham, J. Peter. 2010. "Putting Somali Piracy in Context." *Journal of Contemporary African Studies* 28 (3): 325–41. https://doi.org/10.1080/02589001.2010.499233.
Phandeeyar, Mido, Burma Monitor, Center for Social Integrity, Equality Myanmar, and Myanmar Human Rights Education Network. 2018. "Open Letter to Mark Zuckerberg." April 5, 2018. https://assets.documentcloud.org/documents/4432469/Myanmar-Open-Letter-to-Mark-Zuckerberg.pdf.
Post, David G. 1995. "Anarchy State and the Internet." *Journal of Online Law*, *Art.* 3.
Postema, Gerald J. 1982. "Coordination and Convention at the Foundations of Law." *The Journal of Legal Studies* 11 (1): 165–203. https://doi.org/10.1086/467697.
Pottinger, Nicole E. 2018–2019. "Don't Forget to Subscribe: Regulation of Online Advertising Evaluated through YouTube's Monetization Problem." *Kentucky Law Journal* 107 (3): 515–46.
Purnell, Newley, and Jeff Horwitz. 2020. "Facebook's Hate-Speech Rules Collide with Indian Politics." *Wall Street Journal*, August 14, 2020, sec. Tech. www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346.
Radu, Roxana, Nicolo Zingales, and Enrico Calandro. 2015. "Crowdsourcing Ideas as an Emerging Form of Multistakeholder Participation in Internet Governance: Crowdsourcing Participation in Internet Governance." *Policy & Internet* 7 (3): 362–82. https://doi.org/10.1002/poi3.99.
Ramsay, James Henry. 1925. *A History of the Revenues of the Kings of England, 1066-1399*. Vol. 1. 2 vols. Oxford: The Clarendon Press.
Rawls, John. 1985. "Justice as Fairness: Political Not Metaphysical." *Philosophy & Public Affairs* 14 (3): 223–51.
1999. *A Theory of Justice. Rev*. Cambridge, MA: Belknap Press of Harvard University Press.

Raymond, Mark, and Laura DeNardis. 2015. "Multistakeholderism: Anatomy of an Inchoate Global Institution." *International Theory* 7 (3): 572–616. https://doi.org/10.1017/S1752971915000081.

Reich, Rob, Mehran Sahami, and Jeremy M Weinstein. 2021. *System Error: Where Big Tech Went Wrong and How We Can Reboot*. London, UK: Hodder & Stoughton.

Reidenberg, Joel R. 1998. "Lex Informatica: The Formulation of Information Policy Rules Through Technology." *Texas Law Review* 76 (3): 553–93.

Riedl, Martin J., Sharon Strover, Tiancheng Cao, Jaewon R. Choi, Brad Limov, and Mackenzie Schnell. 2021. "Reverse-Engineering Political Protest: The Russian Internet Research Agency in the Heart of Texas." *Information, Communication & Society* 25 (15): 2299–316. https://doi.org/10.1080/1369118X.2021.1934066.

Riker, William H. 1982. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. Long Grove, IL: Waveland Press.

Ripstein, Arthur. 2004. "Authority and Coercion." *Philosophy & Public Affairs* 32 (1): 2–35.

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

Robertson, Adi. 2018. "Mark Zuckerberg Wants to Democratize Facebook — Here's What Happened When He Tried." *The Verge*, April 5, 2018. www.theverge.com/2018/4/5/17176834/mark-zuckerberg-facebook-democracy-governance-vote-failure.

Robinson, Greg, and Toni Robinson. 2005. "Korematsu and Beyond: Japanese Americans and the Origins of Strict Scrutiny Judgments Judged and Wrongs Remembered: Examining the Japanese American Civil Liberties Cases on Their Sixtieth Anniversary." *Law and Contemporary Problems* 68 (2): 29–56.

Rodriguez, Katitza. 2021. "Indonesia's Proposed Online Intermediary Regulation May Be the Most Repressive Yet." *Electronic Frontier Foundation Deeplinks* (blog). February 16, 2021. www.eff.org/deeplinks/2021/02/indonesias-proposed-online-intermediary-regulation-may-be-most-repressive-yet.

Romm, Tony, and Rachel Lerman. 2021. "Amazon Suspends Parler, Taking Pro-Trump Site Offline Indefinitely." *Washington Post*, January 11, 2021. www.washingtonpost.com/technology/2021/01/09/amazon-parler-suspension/.

Roose, Kevin, Mike Isaac, and Sheera Frenkel. 2020. "Facebook Struggles to Balance Civility and Growth." *The New York Times*, November 24, 2020, sec. Technology. www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html.

Rosenthal, Brian M. 2019. "'They Were Conned': How Reckless Loans Devastated a Generation of Taxi Drivers." *The New York Times*, May 19, 2019, sec. New York. www.nytimes.com/2019/05/19/nyregion/nyc-taxis-medallions-suicides.html.

Rozenshtein, Alan Z. 2021. "Silicon Valley's Speech: Technology Giants and the Deregulatory First Amendment." *Journal of Free Speech Law* 1 (Symposium Issue): 337–76.

Ruhl, J.B. 2011. "General Design Principles for Resilience and Adaptive Capacity in Legal Systems—With Applications to Climate Change Adaptation." *North Carolina Law Review* 89 (5): 1373–403.

Samuelson, Pamela. 2021. "The Push for Stricter Rules for Internet Platforms." *Communications of the ACM* 64 (3): 26–28. https://doi.org/10.1145/3447251.

Sances, Michael W., and Hye Young You. 2017. "Who Pays for Government? Descriptive Representation and Exploitative Revenue Sources." *The Journal of Politics* 79 (3): 1090–94. https://doi.org/10.1086/691354.

Sarno, David. 2009. "Facebook Governance Vote Is a Homework Assignment No One Did." *Los Angeles Times Technology Blog (blog)*. April 23, 2009. www.latimes.com/archives/blogs/technology-blog/story/2009-04-23/facebook-governance-vote-is-a-homework-assignment-no-one-did.

Satariano, Adam, and Mike Isaac. 2021. "The Silent Partner Cleaning Up Facebook for $500 Million a Year." *The New York Times*, August 31, 2021, sec. Technology. www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html.

Satz, Debra. 2010. *Why Some Things Should Not Be For Sale: The Moral Limits of Markets*. New York: Oxford University Press.

Sánchez-Cuenca, Ignacio. 2017. "From a Deficit of Democracy to a Technocratic Order: The Postcrisis Debate on Europe." *Annual Review of Political Science* 20 (1): 351–69. https://doi.org/10.1146/annurev-polisci-061915-110623.

Schauer, Frederick. 1991. "Rules and the Rule of Law." *Harvard Journal of Law and Public Policy* 14 (3): 645–94.

Schechner, Sam. 2021. "Facebook Is Rebuked by Oversight Board over Transparency on Treatment of Prominent Users." *Wall Street Journal*, October 21, 2021, sec. Tech. www.wsj.com/articles/facebooks-oversight-board-says-company-wasnt-fully-forthcoming-on-treatment-of-high-profile-users-11634817601.

Schiffer, Zoe. 2021. "Watch a Police Officer Admit to Playing Taylor Swift to Keep a Video Off YouTube." *The Verge*, July 1, 2021. www.theverge.com/2021/7/1/22558292/police-officer-video-taylor-swift-youtube-copyright.

Schmitt, Carl. 1996. *The Concept of the Political*. Chicago: University of Chicago Press.

2005. *Political Theology: Four Chapters on the Concept of Sovereignty*. Translated by George Schwab. University of Chicago Press ed. Chicago: University of Chicago Press.

2014. *Dictatorship: From the Origin of the Modern Concept of Sovereignty to Proletarian Class Struggle*. Translated by Michael Hoelzl and Graham Ward. English edition. Cambridge, UK; Malden, MA: Polity Press.

Schreckinger, Ben. 2022. "Elon Musk's Release of Twitter Documents on Hunter Biden Has Slowed. Here's Why." *Politico*, December 8, 2022. www.politico.com/news/2022/12/08/twitter-files-hunter-biden-laptop-00072919.

Schulz, Wolfgang. 2022. "Changing the Normative Order of Social Media from Within: Supervisory Bodies." In *Constitutionalising social media*, edited by Edoardo Celeste, Amélie Heldt, and Clara Iglesias Keller, 237–50. New York: Bloomsbury.

Schwarz, Jonas Andersson. 2017. "Platform Logic: An Interdisciplinary Approach to the Platform-Based Economy." *Policy & Internet* 9 (4): 374–94. https://doi.org/10.1002/poi3.159.

Scola, Nancy, and Mohana Ravindranth. 2021. "Biden Looks Past Anger at Silicon Valley to Get Help on Vaccines." *Politico*, February 16, 2021. www.politico.com/news/2021/02/16/biden-silicon-valley-vaccine-help-468995.

Scott, James C. 2008. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed. Yale Agrarian Studies*. New Haven, CT: Yale University Press.

Scott, Mark, and Mike Isaac. 2016. "Facebook Restores Iconic Vietnam War Photo It Censored for Nudity." *The New York Times*, September 9, 2016, sec. Technology. www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html.

Scott, W. Richard. 2004. "Reflections on a Half-Century of Organizational Sociology." *Annual Review of Sociology* 30 (1): 1–21. https://doi.org/10.1146/annurev.soc.30.012703.110644.

Seetharaman, Deepa. 2018. "Facebook's Lonely Conservative Takes on a Power Position." *Wall Street Journal*, December 23, 2018, sec. Tech. www.wsj.com/articles/facebooks-lonely-conservative-takes-on-a-power-position-11545570000.

Seitz, Amanda. 2021. "America 'on Fire': Facebook Watched as Trump Ignited Hate." *Associated Press*, October 28, 2021. https://apnews.com/article/the-facebook-papers-trump-george-floyd-hate-speech-violence-b0f6f26f3fdf889c090703cc2fa8dce0.

Shaw, Aaron, and Benjamin M. Hill. 2014. "Laboratories of Oligarchy? How the Iron Law Extends to Peer Production." *Journal of Communication* 64 (2): 215–38. https://doi.org/10.1111/jcom.12082.

Siapera, Eugenia, and Paloma Viejo-Otero. 2021. "Governing Hate: Facebook and Digital Racism." *Television & New Media* 22 (2): 112–30. https://doi.org/10.1177/1527476420982232.

Silverman, Craig, and Ryan Mac. 2021. "New York Times Columnist David Brooks Wrote a Blog Post For Facebook's Corporate Website." *BuzzFeed News*. February 26, 2021. www.buzzfeednews.com/article/craigsilverman/nyt-david-brooks-blogs-for-facebook.

Silverman, Craig, Craig Timberg, Jeff Kao, and Jeremy B. Merrill. 2022. "Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan. 6 Attack, Records Show." *ProPublica*, January 4, 2022. www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show.

Simon, William H. 2010. "New Governance Anxieties: A Deweyan Response." *Wisconsin Law Review* 2010 (2): 727–36.

Slemrod, Joel, and John D. Wilson. 2009. "Tax Competition with Parasitic Tax Havens." *Journal of Public Economics* 93 (11-12): 1261–70. https://doi.org/10.1016/j.jpubeco.2009.08.004.

Smith, Matthew. 2020 "Facebook Wanted to Be a Force for Good in Myanmar. Now It Is Rejecting a Request to Help with a Genocide Investigation." *Time*. Accessed May 6, 2021. https://time.com/5880118/myanmar-rohingya-genocide-facebook-gambia/.

Solon, Olivia. 2017. "Google's Bad Week: YouTube Loses Millions as Advertising Row Reaches US." *The Guardian*, March 25, 2017. www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon.

2020. "Sensitive to Claims of Bias, Facebook Relaxed Misinformation Rules for Conservative Pages." *NBC News*. August 7, 2020. www.nbcnews.com/tech/tech-news/sensitive-claims-bias-facebook-relaxed-misinformation-rules-conservative-pages-n1236182.

Spangler, Todd. 2021. "OnlyFans Drops Planned Porn Ban, Will Continue to Allow Sexually Explicit Content." *Variety* (blog). August 25, 2021. https://variety.com/2021/digital/news/onlyfans-drops-porn-ban-sexually-explicit-policy-1235048705/.

Srivastava, Swati. 2021. "Algorithmic Governance and the International Politics of Big Tech." *Perspectives on Politics*, November 1–12, 2021. https://doi.org/10.1017/S1537592721003145.

2022. *Hybrid Sovereignty in World Politics. Cambridge Studies in* International Relations *160*. Cambridge, United Kingdom; New York, NY: Cambridge University Press.

Srnicek, Nick. 2016. *Platform Capitalism*. Theory Redux. Cambridge, UK; Malden, MA: Polity.

Stamos, Alex. 2018. "How Does Facebook Investigate Cyber Threats and Information Operations?" *Meta Newsroom (blog)*. November 14, 2018. https://about.fb.com/news/2018/11/investigating-threats/.

Stanford Global Digital Policy Incubator. 2019. "Social Media Councils: From Concept to Reality." Stanford Global Digital Policy Incubator Conference Report.

Stecklow, Steve. 2018. "Why Facebook Is Losing the War on Hate Speech in Myanmar." *Reuters*, August 15, 2018. www.reuters.com/investigates/special-report/myanmar-facebook-hate/.

Steffek, Jens, Claudia Kissling, and Patrizia Nanz, eds. 2008. *Civil Society Participation in European and Global Governance: A Cure for the Democratic Deficit?* Transformations of the State. Basingstoke [England]; New York: Palgrave Macmillan.

Stokel-Walker, Chris. 2020. "What It's Like to Get Locked Out of Google Indefinitely." *Business Insider*, October 31, 2020. www.businessinsider.com/google-users-locked-out-after-years-2020-10.

Stone, Brad. 2009. "Amazon Erases Orwell Books From Kindle." *The New York Times*, July 18, 2009, sec. Technology. www.nytimes.com/2009/07/18/technology/companies/18amazon.html.

Streckfuss, David. 1995. "Kings in the Age of Nations: The Paradox of Lèse-Majesté as Political Crime in Thailand." *Comparative Studies in Society and History* 37 (3): 445–75. https://doi.org/10.1017/S0010417500019769.

Subramanian, Samanth. 2017. "Meet the Macedonian Teens Who Mastered Fake News and Corrupted the US Election." *Wired*, February 15, 2017. www.wired.com/2017/02/veles-macedonia-fake-news/.

Sullivan, Becky. 2021. "Kyle Rittenhouse Is Acquitted of All Charges in the Trial over Killing 2 in Kenosha." *NPR*, November 19, 2021, sec. Law. www.npr.org/2021/11/19/1057288807/kyle-rittenhouse-acquitted-all-charges-verdict.

Sullivan, Mark. 2019. "Exclusive: The Harvard Professor behind Facebook's Oversight Board Defends Its Role." *Fast Company*, July 8, 2019. www.fastcompany.com/90373102/exclusive-the-harvard-professor-behind-facebooks-oversight-board-defends-its-role.

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

Suzor, Nicolas P., Sarah Myers West, Andrew Quodling, and Jillian York. 2019. "What Do We Mean When We Talk about Transparency? Toward Meaningful Transparency in Commercial Content Moderation." *International Journal of Communication* 13: 1526–43.

Swan, Jonathan. 2020. "Trump Privately Scolded, Warned by Allies over Violent Protest Rhetoric." *Axios*, May 31, 2020. www.axios.com/2020/05/31/trump-protests-riots.

Swann, Thomas. 2018. "Towards an Anarchist Cybernetics: Stafford Beer, Self-Organisation and Radical Social Movements." *Ephemera: Theory & Politics in Organization* 18 (3): 427–56.

Teachout, Zephyr. 2020. *Break 'Em up: Recovering Our Freedom from Big Ag, Big Tech, and Big Money*. First edition. New York: All Points Books.

Thomas, June Manning, and Julia Darnton. 2006. "Social Diversity and Economic Development in the Metropolis." *Journal of Planning Literature* 21 (2): 153–68. https://doi.org/10.1177/0885412206292259.

Thompson, Edward P. 1975. *Whigs and Hunters: The Origin of the Black Act*. London: Allen Lane.

Tiku, Nitasha. 2020. "A Top Google Exec Pushed the Company to Commit to Human Rights. Then Google Pushed Him Out, He Says." *Washington Post*, January 2, 2020. www.washingtonpost.com/technology/2020/01/02/top-google-exec-pushed-company-commit-human-rights-then-google-pushed-him-out-he-says/.

Timberg, Craig. 2021. "New Whistleblower Claims Facebook Allowed Hate, Illegal Activity to Go Unchecked." *Washington Post*, October 22, 2021. www.washingtonpost.com/technology/2021/10/22/facebook-new-whistleblower-complaint/.

Timberg, Craig, Elizabeth Dwoskin, and Reed Albergotti. 2021. "Inside Facebook, Jan. 6 Violence Fueled Anger, Regret over Missed Warning Signs." *Washington Post*, October 22, 2021, sec. Technology. www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/.

Titheradge, Noel. 2021. "OnlyFans: How It Handles Illegal Sex Videos – BBC Investigation." *BBC News*, August 19, 2021, sec. UK. www.bbc.com/news/uk-58255865.

Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–40. https://doi.org/10.1017/S0020818307070282.

Traill, John S. 1975. "The Political Organization of Attica: A Study of the Demes, Trittyes, and Phylai, and Their Representation in the Athenian Council." *Hesperia Supplements* 14: i–169. https://doi.org/10.2307/1353928.

Tromholt, Morten. 2016. "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being." *Cyberpsychology, Behavior, and Social Networking* 19 (11): 661–66. https://doi.org/10.1089/cyber.2016.0259.

Tucker, Joshua A., et al. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." *ID 3144139*. Rochester, NY: Hewlett Foundation. www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

Tussey, Deborah. 2014. "Facebook: The New Town Square." *Southwestern Law Review* 44 (2): 385–410.

Tworek, Heidi. 2019. *Social Media Councils*. Waterloo, ON: Centre for International Governance Innovation. www.cigionline.org/articles/social-media-councils/.

Tworek, Heidi, Ronan Ó Fathaigh, Lisanne Bruggeman, and Chris Tenove. 2020. "Dispute Resolution and Content Moderation: Fair, Accountable, Independent, Transparent, and Effective." *Translatlantic Working Group*. www.ivir.nl/publicaties/download/Dispute_Resolution_Content_Moderation_Final.pdf.

Tyler, Tom R. 1990. *Why People Obey the Law*. New Haven: Yale University Press.

1997. "Procedural Fairness and Compliance with the Law." *Swiss Journal of Economics and Statistics* 133 (2): 219–40.

2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57 (1): 375–400. https://doi.org/10.1146/annurev.psych.57.102904.190038.

2009. "Legitimacy and Criminal Justice: The Benefits of Self-Regulation." *Ohio State Journal of Criminal Law* 7 (1): 307–60.

U.S. Immigration and Customs Enforcement. 2020. "IPR Center, Amazon Launch 'Operation Fulfilled Action' to Stop Counterfeits." *ICE Newsroom (blog)*. November 24, 2020. www.ice.gov/news/releases/ipr-center-amazon-launch-operation-fulfilled-action-stop-counterfeits.

United Nations. 2008. "Arab Rights Charter Deviates from International Standards, Says UN Official." *UN News*. January 30, 2008. https://web.archive.org/web/20220120164141/https://news.un.org/en/story/2008/01/247292-arab-rights-charter-deviates-international-standards-says-un-official.

United Nations Human Rights Council. 2018. "Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar." A/HRC/39/CRP.2. www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_CRP.2.pdf.

Van der Spuy, Anri. 2017. *What if We All Governed the Internet?: Advancing Multistakeholder Participation in Internet Governance*. Paris: UNESCO Publishing.

Van Loo, Rory. 2016. "The Corporation as Courthouse." *Yale Journal on Regulation* 33 (2): 547–602.

2021. "Federal Rules of Platform Procedure." *University of Chicago Law Review* 88 (4): 829–94.

VanLandingham, Rachel E. 2017. "Jailing the Twitter Bird: Social Media, Material Support to Terrorism, and Muzzling the Modern Press." *Cardozo Law Review* 39 (1): 1–55.

Vlamis, Kelsey. 2022. "Twitter's Former Trust and Safety Chief Said It Was a Mistake to Censor the Hunter Biden Laptop Story: 'We Didn't Know What to Believe'." *Business Insider*, November 30, 2022. www.businessinsider.com/yoel-roth-twitter-censor-hunter-biden-laptop-story-was-mistake-2022-11.

Wagner, Kurt. 2021. "Facebook Says It Has Spent $13 Billion on Safety and Security Efforts Since 2016." *Fortune*, September 21, 2021. https://fortune.com/2021/09/21/facebook-says-it-has-spent-13-billion-on-safety-and-security-efforts-since-2016/.

Wakabayashi, Daisuke, and Scott Shane. 2018. "Google Will Not Renew Pentagon Contract That Upset Employees." *The New York Times*, June 1, 2018, sec. Technology. www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html.

Waldman, Ari Ezra. 2021. *Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power*. Cambridge, United Kingdom; New York, NY, USA: Cambridge University Press.

Warofka, Alex. 2018. "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar." *About Facebook (blog)*. November 6, 2018. https://about.fb.com/news/2018/11/myanmar-hria/.

Weber, Max. 1946. *From Max Weber: Essays in Sociology*. Edited by Hans Heinrich Gerth and Charles Wright Mills. New York: Oxford University Press.

Weingast, Barry R., and William J. Marshall. 1988. "The Industrial Organization of Congress; or, Why Legislatures, Like Firms, Are Not Organized as Markets." *Journal of Political Economy* 96 (1): 132–63. https://doi.org/10.1086/261528.

Weiser, Philip J. 2009. "The Future of Internet Regulation." *UC Davis Law Review* 43: 529–90.

Wennerberg, Hjalmar. 1967. "The Concept of Family Resemblance in Wittgenstein's Later Philosophy." *Theoria* 33 (2): 107–32. https://doi.org/10.1111/j.1755-2567.1967.tb00614.x.

Westley, Frances. 1995. "Governing Design: The Management of Social Systems and Ecosystems Management." In *Barriers and Bridges to the Renewal of Ecosystems and Institutions*, edited by Lance H. Gunderson, Crawford S. Holling, and Stephen S. Light, 391–427. New York: Columbia University Press.

Whelan, Frederick G. 1983. "Prologue: Democratic Theory and the Boundary Problem." In *Nomos 25: Liberal Democracy*, edited by James Roland Pennock and John W. Chapman, 13–47. New York: New York University Press.

Williamson, Oliver E. 1996. "Economics and Organization: A Primer." *California Management Review* 38 (2): 131–46. https://doi.org/10.2307/41165836.

2005. "Why Law, Economics and Organization?" *Annual Review of Law and Social Science* 1 (1): 369–96. https://doi.org/10.1146/annurev.lawsocsci.1.031805.111122.

Wilson, Richard Ashby. 2015. "Inciting Genocide with Words." *Michigan Journal of International Law* 36 (2): 277–320.

Wilson, Richard Ashby, and Molly K. Land. 2021. "Hate Speech on Social Media: Towards a Context-Specific Content Moderation Policy." *Connecticut Law Review* 52 (3): 1029–76.

Wofford, Benjamin. 2022. "The Infinite Reach of Joel Kaplan, Facebook's Man in Washington." *Wired*, March 10, 2022. www.wired.com/story/facebook-joel-kaplan-washington-political-influence/.

Wu, Tim. 2018. *The Curse of Bigness: Antitrust in the New Gilded Age*. New York, NY: Columbia Global Reports.

Wymant, Chris, et al. 2021. "The Epidemiological Impact of the NHS COVID-19 App." *Nature* 594 (7863): 408–12. https://doi.org/10.1038/s41586-021-03606-z.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *The Quarterly Journal of Economics* 129 (4): 1947–94. https://doi.org/10.1093/qje/qju020.

Young, Iris Marion. 2001. "Activist Challenges to Deliberative Democracy." *Political Theory* 29 (5): 670–90. https://doi.org/10.1177/0090591701029005004.

Young, Oran R. 2002. "Institutional Interplay: The Environmental Consequences of Cross-Scale Interactions." In *The Drama of the Commons*, edited by Elinor Ostrom, Thomas Dietz, Nives Dolšak, Paul C. Stern, Susan Stonich, and Elke U. Weber, 263–91. Washington, DC: National Academy Press.

Yu, Peter K. 2006. "TRIPS and Its Discontents." *Marquette Intellectual Property Law Review* 10 (2): 369–410.

Zakrzewski, Cat, Gerrit De Vynck, Niha Masih, and Shibani Mahtani. 2021. "How Facebook Neglected the Rest of the World, Fueling Hate Speech and Violence in India." *Washington Post*, October 24, 2021, sec. Tech Policy. www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/.

Zannettou, Savvas, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. 2018. "What Is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber." In Companion *of The Web Conference 2018 on The Web Conference 2018 – WWW '18*, 1007–14. Lyon, France: ACM Press. https://doi.org/10.1145/3184558.3191531.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12 (1): 415–38. https://doi.org/10.1146/annurev-economics-081919-050239.

Ziady, Hanna. 2019. "Twitter Explains Why It Hasn't Suspended Trump's Account." *CNN Business*, October 16, 2019. www.cnn.com/2019/10/16/tech/twitter-trump-rules/index.html.

Zittrain, Jonathan. 9-24-21. "The Inexorable Push For Infrastructure Moderation." *Techdirt.* (blog). 9-24-21. www.techdirt.com/articles/20210924/12012347622/inexorable-push-infrastructure-moderation.shtml.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs.

Zurth, Patrick. 2021. "The German NetzDG as Role Model or Cautionary Tale? – Implications for the Debate on Social Media Liability." *Fordham Intellectual Property, Media & Entertainment Law Journal* 31 (4): 1084–1153. https://doi.org/10.2139/ssrn.3668804.

Zuylen-Wood, Simon van. 2019. "'Men Are Scum': Inside Facebook's War on Hate Speech." *Vanity Fair*, February 26, 2019. www.vanityfair.com/news/2019/02/men-are-scum-inside-facebook-war-on-hate-speech.

Index

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press

https://doi.org/10.1017/9781108975438 Published online by Cambridge University Press