12

Platform Power, Online Speech, and the Search for New Constitutional Categories

NATHANIEL PERSILY

What position should large social media platforms occupy in American constitutional law?

Are they like political parties[1] or one-company towns?[2] In that case, they would be state actors that must respect the First Amendment rights of their users. Or are they like railroads,[3] telephone companies,[4] or Federal Express[5]—common carriers that ordinarily cannot discriminate among their customers in delivering goods or information from place to place? Or perhaps, as Justice Clarence Thomas suggested in a recent opinion,[6] they are like "places of public accommodation,"[7] akin to hotels and restaurants that are regulated by antidiscrimination laws? Or are they like a shopping mall[8] that must allow protesters to hand out leaflets in its parking lot, or a law school[9] that must allow the military to recruit students on campus? Or are they public utilities, like an electric company,[10] which enjoy First Amendment rights but also can be subject to considerable regulation? Or are they like a broadcast television station,[11] a cable provider,[12] or a newspaper?[13] If they are like one of those, the government may or may not have the power to require them to carry diverse and contrary points of view. Or are they simply like other corporations with First Amendment rights against government speech regulation, even including the corporation's right to spend unlimited amounts of money seeking to influence election campaigns?[14]

All such analogies have something to offer, but none fits perfectly. The internet and the rise of social media require the creation of new legal categories even as we struggle to fit these new institutions and relationships into old conceptual boxes. That each analogy to an earlier form of expression may fall short should not surprise us: Each generation tends to view technologies of its day as

unprecedented and in need of new legal definitions. The question that these new platforms pose, however, is: How can we create a legal environment that protects individuals' ability to reach an audience on the internet while also recognizing the twin dangers of unfettered expression that leads to real-world harm and unconstrained platform power that allows for consolidated corporate domination of the speech marketplace?

While we struggle to characterize the nature of social media platforms under the Constitution, courts and even the platforms themselves have described the environment these platforms have created as "the new public square."[15] As the Court majority put it in *Packingham v. North Carolina*, "[t]hey allow a person with an Internet connection to 'become a town crier with a voice that resonates farther than it could from any soapbox.'"[16] The Court was considering a law that restricted access to social media, not a practice by a social media company regulating content on its platform. However, even the leaders of Facebook and Twitter have joined in the chorus to characterize their platform environments in the same terms.[17] The "public square" metaphor has had predictable consequences, leading some to argue for increased regulation of social media companies given the outsized importance of the speech domain they control and others to argue in favor of treating their own self-regulation as akin to government regulation of individual speech.[18] Facebook has as much power as any government (and more than most), the argument goes, so we should consider its regulations of hate speech, incitement, and so on, as we would those of the government. Similarly, if the Facebook Newsfeed is now the public square, any regime of speech regulation (e.g., political advertising laws) which leaves out social media is ignoring the principal arena in which speech (dangerous or otherwise) is taking place.

Finding the right legal category for new social media platforms will not dictate how they should be regulated or how they should behave if unregulated. However, it represents a first step toward defining a permissible range of regulatory outcomes. As we explore different models, it is important to account for the unique features of social media platforms, in general, or even a given platform, in particular. Assuming we can identify the universe of "large social media platforms" to which these new rules should apply (itself a challenging task), these new rules ought to account for the (1) revolutionary speed of communication on the internet, (2) the need for automated filtering that operate as prior restraints, and (3) the global scope of these platforms. They also must recognize the distinctive characteristic of these social media platforms—one that makes them different from the traditional public square—namely, (4) the algorithms that organize and prioritize communication by making judgments that are inevitably content- and viewpoint-based.

This essay first analyzes and rejects the arguments that internet platforms are state actors, common carriers, or places of public accommodation, in the traditional sense. The next section explains why neither First Amendment law nor international human rights law—developed to deal with government speech suppression—has much to offer when it comes to guiding content moderation decisions. To be sure, case decisions in whatever forum dealing with general issues of free speech can be helpful (just as a good law review article would not be irrelevant), but there are critical differences between the law developed in these contexts and the analysis needed to deal with platform content regulation. The final section explains why and what to do about it. Certainly, we need to be worried about the power these private monopolies have over the speech marketplace. But different rules should apply to different products, and platforms should have a wide degree of latitude to impose the types of content- and viewpoint-based regulations that would be off-limits to government. The Conclusion discusses these issues in light of recent disclosures made by the Facebook whistleblower Frances Haugen and recent actions taken in Europe to regulate large online platforms.

This essay oscillates between positive arguments about the nature of platforms (constitutionally or jurisprudentially speaking) and normative arguments about how governments should regulate them and how platforms should regulate speech. To be clear from the outset, my view is that very few rules that have bound government speech regulation should bind platforms. This means that (1) the First Amendment (and international human rights law applicable to governments) does not and should not constrain platform content moderation; (2) the government should *not* pass a statute that requires those same legal principles to be applied by platforms on their own (as the state of Florida has and some in Congress have urged); and (3) the platforms, even absent government compulsion, should not, on their own, constrain themselves according to those principles, even if they have a "right" to do so. This may seem, at first, like the most extreme version of an argument for platform autonomy. However, at the same time as I would urge rejection of the platforms-as-governments analogy, I would also give very little credence to the platforms-as-speakers argument—that is, the notion that the government cannot and should not regulate large platforms for fear of violating the platforms' First Amendment rights. Indeed, neither the Constitution nor some other abstract theory related to constraints of government power over corporate speakers should stand in the way of government regulation of "large online platforms" in the area of content moderation or in the fields of antitrust, privacy, transparency, advertising, and taxation. Indeed, regulation in these putatively nonspeech arenas may go a long way toward addressing

some of the concerns about corporate power that undergird the worries about the damage platforms cause democracy through their content moderation policies. Government regulation of social media might implicate the constitutional rights of users, but the platforms themselves (contra *Citizens United*) should not be viewed as having rights that stand in the way of regulation of content moderation or other practices.

One final prefatory note as to the universe of actors considered in this analysis: One of the problems with commentary related to the dangers of "big tech" and social media is that several different types of communication platforms with very different "affordances" are grouped together as if their content moderation or free speech impacts are the same. However, as scholars such as evelyn douek have observed, content moderation questions affect everything from video games to restaurant reviews to home exercise equipment like Peloton.[19] Anytime a platform allows a user to "speak" to other users, the platform must come up with rules about what is permissible speech.

Moreover, some "platforms" comprise many different products. Not only does Facebook own Instagram, WhatsApp, and Oculus, but even the Facebook platform itself can include Messenger (which is also available as an independent application) and other features. Similarly, Google Search presents very different speech questions than do Gmail, Google Chat, YouTube, or the Android App Store. We should not expect the same rules of content moderation (as prescribed by governments or platforms) to apply to every mode of communication that a platform offers. WhatsApp is basically a glorified text messaging service, but the fact that Facebook owns WhatsApp probably does not mean that that messaging service should be subject to rules different than those that govern Apple's iMessage or even text services enabled by a cell phone provider.

Nor should we expect the same rules to apply to every layer of the internet stack. In other words, social media companies may have different obligations than hosts (e.g., Amazon Web Services, DreamHost), transit providers (e.g., Level(3), NTT), reverse proxies/CDNs (e.g., Akamai, Cloudflare), authoritative DNS providers (e.g., Dyn, Cloudflare), registrars (e.g., GoDaddy, Tucows), registries (e.g., Verisign, Afilias), internet service providers (e.g., Comcast, AT&T), recursive DNS providers (e.g., OpenDNS, Google), and browsers.[20] Each of these types of companies must decide what types of "speakers" can use its services. Each also has the potential to "silence" or amplify certain voices.

Finally, even if we can define the class of social media applications to which certain speech rules should apply, we inevitably must distinguish, on a somewhat arbitrary basis, which are large enough to warrant the special rules necessary to protect democracy and the speech marketplace. Facebook and Twitter should be subject to different rules than an upstart like Gettr or Parler, let alone the comment sections of a website like the *New York Times*, Fox News, Breitbart, or DailyKos. But what about TikTok, Reddit, Clubhouse, Next Door, or Discord?

Should the rules (again, either as to the propriety or constitutionality of government regulation or the rules the platform itself should enact) depend on the size of the user base, the professed openness of the platform, or some other metric associated with the danger the platform poses? At a certain point, the speech discussion begins to bleed over into a similar discussion occurring among antitrust experts concerning platform power—that is, once a platform achieves certain scale due to network effects it should be subject to greater regulation by government and its content moderation rules take on greater significance. For purposes of this essay, though, assume that the universe of relevant actors includes (at least) the Facebook newsfeed, YouTube, Twitter, and TikTok, based simply and arbitrarily on the size of the US user base, which for all of these is in excess of fifty million monthly active users.

Platforms as State Actors, Common Carriers, Public Accommodations, and Essential Facilities

State Action

If large internet companies are state actors, then they have no rights to object to government regulation, and they must respect the First Amendment and other rights of their users. Companies can become state actors if they perform a function that has been "traditionally, exclusively reserved to the state"[21] or if they are so intertwined with the state (through funding or joint performance of state-directed policy) that they are, in effect, agents of the state. Given the often adversarial relationship of the platforms to the government, it would be difficult to consider such corporations state actors in all their functions and incarnations. That a firm is powerful—even a powerful monopoly—does not mean it suddenly converts into the state. More is necessary. We can fear the power of Facebook and Google and even conclude they should be regulated into oblivion, but they are still private companies unless certain other factors are present. State action depends not merely on the size and importance of the company but also on the type of power it exercises and its proximity to formal organs of the state or state officials.

In the class action complaints he filed against Facebook, Google, and Twitter, following his deplatforming in the wake of the attempted insurrection on January 6, 2021, President Trump argued that the companies are state actors violating his First Amendment rights.[22] The suit is frivolous, but it raises important questions about what, if anything, might convert these companies into institutions that must obey the First Amendment. Although, like so many politicians and pundits do when they attack the platforms for viewpoint discrimination, the complaint emphasizes the platform's scale and the "public square" argument from *Packingham* and elsewhere,[23] it also grasps for other more traditional state action

arguments. It tries to latch onto the series of cases related to government entanglement with private actors, to suggest that the federal government encouraged, coerced, or partnered with these platforms to silence his and others' speech.[24]

Specifically, the complaint argues: "In censoring the specific speech at issue in this lawsuit and deplatforming plaintiff, defendants were acting in concert with federal officials, including officials at the CDC and the Biden transition team."[25] Whether through a series of implied threats from Democrats (what Genevieve Lakier discusses more generally as "jawboning"),[26] partnership with the government through free advertising, or reliance on agencies as official sources of information against which private posts would be judged as disinformation, the platforms, under this view, have become handmaidens in a government censorship program. They are all the more state-like, the complaint maintains, because Section 230 of the Communications Decency Act (which, they also argue, happens to be unconstitutional) grants platforms the benefits of immunity from liability for user-generated speech that they host and for takedowns of speech deemed "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable."[27] This "benefit" of Section 230 immunity, like generous government funding or using government facilities, the argument goes, converts their content moderation decisions into state action and censorship in violation of the First Amendment.

The lawsuit is absurd for a number of reasons, not the least of which is that the plaintiff in this case happened to be the elected leader of the government alleged to have censored him at the time. Government officials can use their bully pulpits to excoriate internet platforms without turning them into state actors, and they can run advertisements or even give money to corporations (whether internet platforms, TV networks, hospitals, private schools, or restaurants) without those recipients turning into "the government." Indeed, if the government is coercing the platforms, as suggested, then a lawsuit against the United States or perhaps individual officeholders might be appropriate. Moreover, if Section 230 is a sufficient "hook" for state action, then any website that moderates user comments is a state actor, as they all benefit from the immunities that the statute provides. Section 230 is not limited to "big tech" or monopolistic websites, as impoverished popular commentary suggests. It applies to all websites that allow for user-generated content. If mere protection by Section 230 creates state actors, then much of the web is "the government."[28]

Common Carriers or Public Accommodations

A more serious analysis of the constitutional position of large internet platforms suggests that they be treated akin to common carriers (telephones or railroads) or places of public accommodation (hotels and restaurants).[29] There are strong

and weak forms of these arguments. The strong form is not dissimilar to the state action argument—suggesting that, as common carriers, the platforms have a constitutional obligation to welcome all comers. Just as a telephone service could not allow only "Democratic" phone calls, so, too, Facebook could not only allow Democrats to have accounts in its service.

The weak form of the common carrier argument maintains that the platforms can be forced to allow all comers onto their service. In other words, the state may force them, in essence, to comply with First Amendment precedent that would constrain a government—for example, to prohibit them from engaging in content- or viewpoint-based discrimination. Just as a common carrier (or place of public accommodation) could be forced to provide equal access to people of different races or political parties, the same obligation could be imposed on the platforms. Of course, were the state to impose such an obligation on the platforms, much of their efforts at policing disinformation (banning groups like QAnon or antivaxxers) or addressing hate speech (banning white supremacists and terrorist sympathizers), let alone garden-variety moderation of bullying or depictions of violence and nudity, might thereby be preempted.

One key question here is whether some right of the platform—either the platform's own First Amendment rights or perhaps its property rights—should prevent government treatment of them as common carriers or places of public accommodation. Do large platforms have a constitutional right to engage in moderation efforts to create an experience for their users that might exclude certain types of objectionable (but constitutionally protected) content?

Of course, *some* platforms must have such a right. The First Amendment would certainly protect me in my efforts to start a chat group with my friends or those of the Republican Party (or a conservative publication like Breitbart) creating a website limited to Republican users or content. But these are not examples of *common* carriers. Indeed, Facebook itself could not plausibly have been considered a common carrier when Mark Zuckerberg developed it in his Harvard dorm room and limited it to his classmates. The question is whether, once achieving some level of scale or importance for the communication ecosystem, Facebook became a common carrier. Facebook's implicit promises of allowing broad participation on its service irrespective of party, race, and so on might also be important (although that promise is itself cabined by the Community Standards and Terms of Service). The implication here, of course, is that a social media service, which might originally have been conceived as a closed and highly regulated environment, could become a common carrier once it achieves a certain scale and importance.

This is a powerful argument with serious real-world implications, as a recently passed[30] and rapidly enjoined[31] law from Florida highlights. That law imposes significant fines on platforms that deplatform or shadow ban candidates. It also

punishes platforms for taking "any action to censor, deplatform, or shadow ban a journalistic enterprise based on the content of its publication or broadcast."[32] It exempts any platform that operates a theme park (Florida's convenient "Disney exception"), but otherwise applies to any platform with one hundred million global monthly users and brings in one hundred million dollars annually.[33] Florida argued that the platforms were common carriers, making arguments echoing *Pruneyard* and *Rumsfeld*.[34] The court sided with the platforms, though, preliminarily enjoining the law. It held that the law was a content-based restriction on the platform's own First Amendment rights, failing either strict or intermediate scrutiny. Because the Florida law was so poorly written, riddled with content-based exceptions, and publicly described as an attempt to control liberal tech companies,[35] the court did not feel the need to engage deeply with the common carrier argument.

Justice Clarence Thomas has given a more thoughtful, even if still cursory, judicial treatment of the platform-as-common-carrier or public accommodation argument in his concurrence from a vacatur of the decision that held that President Trump's Twitter account was a public forum.[36] Describing the platforms as "networks within a network," Thomas argues that platforms with "substantial market power" that "hold themselves out as open to the public" might be considered common carriers. Gesturing toward the arguments in the Trump lawsuits arising from the Facebook, Twitter, and YouTube deplatforming, Thomas also finds the benefits bestowed through Section 230 and the specter of government coercion of the platforms as suggestive of their common carrier status. He throws Google Search and Amazon into the mix, as well, suggesting the argument extends well beyond social media.

Essential Facilities

What is missing from these accounts, as well as from a series of *Wall Street Journal* op-eds from Professor Phillip Hamburger,[37] is an appreciation and understanding of the different functions that social media platforms perform. Specifically, most of what they do is not "carrying" in the same way as railroads and telephone companies. They are not merely hosting speech, but organizing it. The most important feature of the platforms is the algorithms they employ to structure a unique "feed" for every individual user.

In what is the most sophisticated treatment of the subject, Professor Eugene Volokh develops a more modest common carrier argument restricted to the particular functions of the platform.[38] He considers the platforms as "common carriers" merely in their "carrying" functions—namely, hosting an account, posting content, and allowing users to follow other accounts. On this view, the state could prevent Facebook from deplatforming President Trump (and anyone else violating its community standards) or from preventing

people from following him. It could also prevent takedowns of content from the platform. If it were truly a *common* carrier, then the platform might be forced to keep on all hate speakers, disinformation purveyors, terrorist or supremacist sympathizers, bullies, self-harm videographers, and any number of other speakers that, for example, could not be discriminated against by phone companies or railroads.

However, when it comes to the presentation of information to users in feeds, Volokh (as well as the court that struck down the law in Florida) emphasizes the First Amendment rights of the platforms. The algorithms that organize what users see are, under this view, like the newspaper editors in *Miami Herald v. Tornillo*[39]: exercising legitimate content-based judgments as to what appears where on the page (or in this case, screen). Facebook and Twitter's Newsfeeds, YouTube's recommendation system, and Google Search would be beyond the reach of these common carrier regulations. When a platform organizes, prioritizes, and packages communication, it "speaks" rather than "carries."

This distinction seems plausible, even if, as I will argue later, it is undesirable. It would also allow for common carrier regulation of messaging apps and email. In other words, the state can require a platform to allow all comers to exist on the platform and communicate directly with a willing audience, but it cannot prevent a platform from discriminating against speakers by limiting the reach or visibility of their posts. The Florida law, which prohibited "shadow banning"—the use of algorithms to deprioritize a speaker's content in users' feeds—would go well beyond this minimalist view of common carriers. Even so, the implication of this weak form of the common carrier argument is that the platforms might not be able to deplatform or delete the posts of purveyors of hate, disinformation, violent imagery, and even advocacy of some offline harm.

If the common carrier argument only extends to the hosting function, then the platforms become indistinguishable from the internet itself. The ability to appear on a site and to have users see your content is indistinguishable from the right to have one's own separate website. The platform's added value comes from its algorithms. Facebook reports that when it demotes content it reduces reach by 80 percent.[40] Similarly, 70 percent of YouTube views come from the recommendation or autoplay features.[41] Merely hosting an account with content on it for users who seek it out does not afford the speakers much more than what they would get from having their own URL. In other words, deeming them common carriers for the accounts they host might protect "freedom of speech," but what these users are really fighting for is "freedom of reach,"[42] namely, the likelihood that their content would land in the feeds of a large audience on the platform.

The Place of Free Speech Law in Platform Regulation and Policy

Should the First Amendment Govern Platform Content Moderation?

The more limited common carrier argument is both too restrictive of platform autonomy to regulate speech and too protective of the First Amendment rights of platforms with respect to their algorithms. Platforms need to take action against harmful but constitutionally protected speech, and governments need to be able to regulate more than the mere ability to host accounts. For the largest platforms with the most influential algorithms, the state should enjoy wide latitude to regulate the ingredients of such algorithms to ensure both fairness and transparency.

Just as the state action argument goes too far, so, too, does the strong common carrier argument that the platforms should use the First Amendment as the guidepost for their content moderation decisions (or be compelled to do so). There are certain absurd implications that should be recognized from the outset. *Citizens United* and the constitutional restrictions on political advertising regulation are cases in point. If Twitter were "the state" and the newsfeed were the "public square," its ban on political ads would be unconstitutional,[43] as would Facebook's recent ban on new political ads being launched in the week before an election.[44] Indeed, many of the restrictions Facebook places on ads would be unconstitutional,[45] including its ban on ads with sensational content[46] or those that lead to "low-quality or disruptive" landing pages.[47] Some of these rules are not dissimilar to what a television or radio station might require, but the argument is common among different types of media entities: The legitimate interests and methods of a company to determine the boundaries of speech-for-purchase are different from those that a government can lay down for all of society.

This argument applies to the prioritization of content, as well as the rules for purchasing audience reach through ads. If one were to apply the First Amendment to platform algorithms, then content- and viewpoint-based prioritization of content would be prohibited (or at least subject to strict scrutiny).[48] Yet, in a very real sense, that is what a social media algorithm is. It makes personalized decisions about what you should see based on a host of characteristics about you (most importantly, your viewing history and friend networks), as well as decisions by the company about what kind of product and experience it wants to deliver to the users. If it could only include content-neutral factors in the algorithm, then only a chronological feed (or something similarly ringing in the spirit of "time, place, and manner" regulation) would be allowed. (This is one of the recommendations made by Frances Haugen, the Facebook whistleblower, in her Senate testimony.)

One consequence of content neutrality for social media algorithms is that they would be easily gamed and overrun by spam. The platforms' most aggressive content moderation programs, by far, concern the takedowns of accounts and content deemed to be spam. For example, a chronological feed basically favors speakers who post as often as possible. If you have one hundred Facebook friends, ninety-nine of whom post one post per day, but one of whom (perhaps due to automation, perhaps due to overcaffeination) posts one thousand posts per day, a chronological feed will favor the information superspreader. Of course, there may be content-neutral ways of dealing with this particular example, but the point remains that, for the algorithm to function, it must make "decisions" about speakers and content quality. It is no wonder, then, that Facebook takes down close to four billion accounts per year, mostly due to its antispam enforcement.[49]

Most spam is generated by automated or anonymous accounts. Anonymity, however, is generally protected under the First Amendment.[50] Were that to be the decision rule applied by platforms, real-names policies (like Facebook's) would not be allowed. To be sure, there are instances when disclosure can be mandated by government—as in, for example, campaign spending or advertisements. But if Facebook were a state, it could not require all organic content to be identified with a speaker. As it stands, some platforms, like Twitter, allow for anonymity and others, like Facebook, try to avoid it. But if the First Amendment were the lodestar for content moderation, anonymity would need to be generally protected. Doing so would seriously threaten the platform's ability to police online harm, foreign election interference, and a host of other problems.

It also remains far from clear what it might mean for an algorithm to "obey the First Amendment." Anyone who interacts with these firms quickly appreciates that there are few, if any, people who fully understand all the components of the Facebook newsfeed algorithm or Google search, for example. These "speech regulations" have now evolved over decades and are basically millions of lines of code, some of which are the product of machine learning that is not amenable to the kind of First Amendment analysis one finds in a judicial proceeding.

As compared to the algorithm, however, the platforms' community standards would be amenable to such an analysis, and they would fail miserably. Trump's complaint against the platforms did have this right: The platforms' rules against hate speech, glorification of violence, dangerous individuals, and most other standards, such as nudity, bullying, suicide, and self-injury, would be deemed unconstitutionally vague and overbroad. Almost any restriction on hate speech,[51] disinformation,[52] and glorification of violence[53] would violate the First Amendment. Even in areas where the government has greater latitude, as with obscenity and indecency, the platforms' rules on nudity go well beyond the boundaries the courts have allowed for government.[54] At times, the platforms' rules are both overly specific and overly broad. Facebook, for example, has a

series of very specific rules when it comes to comparisons between certain racial or religious groups and certain types of animals, which means that (given historical and cultural context) some groups, and only some groups, can be compared to some animals, but not others.[55] At the same time, its hate speech rules prevent users from calling for "social exclusion" of protected classes of people, "which means things like denying access to spaces and social services."[56]

One of the key reasons the platforms' rules are sometimes overly specific and overly broad (and often "unconstitutional") is that many such standards must be applied through automated filtering. A computer will have great difficulty, for example, in distinguishing positive (e.g., group-reaffirming) from negative uses of racial epithets or between satire, news coverage, and harmful speech (e.g., reporting on or mocking a hateful speaker while also restating the hate speech). The platforms simply are not in the position to make judicial-style judgments for each violation of the community standards.

In short, the platforms and their automated filters are *in the business* of prior restraints, which is about as electric a third rail as exists in First Amendment law.[57] Speech is, necessarily, prevented on the platform even before it reaches its audience. Both the scale and the speed of online communication make such filtering—which is inevitably overinclusive and underinclusive of the targeted content category—inevitable. Facebook must "adjudicate" millions of pieces of content every day.[58] No number of human moderators would be able to eyeball every post for community standards violations. Moreover, only automated filtering can respond with the speed necessary to enforce the community standards before the violative content goes viral. As any number of examples—such as the Christchurch massacre livestream[59] or the spread of Election Day disinformation—demonstrates, waiting for judicial or even executive-style enforcement action is a luxury the platforms do not enjoy. Either their systems pick up the offending content and limit its distribution, or it quickly may reach an audience of millions or even hundreds of millions of users.

One final note regarding the relevance of scale and whether the First Amendment can guide platform content moderation: Although in certain circumstances platforms can limit their enforcement to a particular speaker and audience in a specific geographic territory (so-called geofencing), their content moderation rules necessarily apply to a global audience. It would be near impossible to have a different hate speech or obscenity policy for every country in which Twitter, Facebook, and YouTube operate. The First Amendment of the US Constitution, awesome though it may be, does not represent a universal norm as to how to draw lines between permissible and impermissible speech. Most of Facebook's audience—indeed, well over 90 percent—live outside of the United States. If Facebook's Community Standards are to represent the Facebook community, imposing one country's

view of speech seems inconsistent with the nature of the platform, its reach, and its diverse user base. Leaving aside the fact that some countries actually ban some speech that the First Amendment clearly protects (potentially pitting a platform's permissive content guidelines against a country's regulation of speech), the platform's rules need to be sensitive to the charge of American First Amendment imperialism over the world's speech marketplace. Our First Amendment tradition is exceptional in a number of respects (defamation, hate speech, obscenity, and incitement are just a few examples). Universal speech rules for platforms based on the unique American tradition would be in persistent tension with the global nature of the platform.

Should International Human Rights Law Govern Platform Content Moderation?

Given the global reach of the platforms, should international human rights law, perhaps, guide platform content moderation decisions? Indeed, many NGOs and scholars have argued as much.[60] The Facebook Oversight Board, pursuant to the Charter developed by Facebook itself,[61] also has applied international law in its decisions on content moderation. If the Oversight Board is going to act like a global speech court, the argument goes, then it stands to reason that it should adopt international speech norms as guideposts for its decision.

However, for the same reasons that the First Amendment—designed as a limit on government control of speech—serves as a poor guide for content moderation, so, too, international human rights law—designed for the same purpose—fails as well. Even if international law is not generally as speech protective as the US First Amendment, it still provides an answer to a different question than the one platforms are answering in developing their content moderation policies.

Two cases from the Facebook Oversight Board demonstrate the mismatch. The first involved a Brazilian breast cancer examination video on Instagram originally taken down by an automated filter under the platform's nudity policy.[62] Facebook, which owns Instagram, later realized this video probably should be allowed under an applicable exception to its rules, as the board eventually held. However, the case exemplifies the difference between human rights law and rules for content moderation.

Were a government to ban all expression involving nudity, health education or otherwise, it would certainly be violating international human rights standards (let alone the First Amendment). But is the harm really the same when a platform applies an automated filter for nudity? Because it will often be difficult to determine "context" and "intent" on a grand scale in a split second, automated filters that will necessarily overenforce will need to be employed.

Given the importance of trying to prevent indecent material, child sexual imagery, and nonconsensual pornography, is the cost of preventing some legitimate and important nude imagery really one a platform should be forced to bear? It is not as if there is a shortage of places on the internet where one might find nudity. Whereas a government should not be able to eradicate nudity from the speech marketplace, a platform operating globally that takes down or filters out millions of nude images per month ought to be able to decide that a clear rule, easily enforced by automation, might be the best way to preserve the platform environment.

The same point could be made with respect to a case involving borderline COVID disinformation that the Oversight Board ordered reinstated onto the platform.[63] The post in question included a video that indirectly questioned the French government's policy of banning the use of hydroxychloroquine in treating COVID, saying it was a harmless drug with some evidence of success. It was removed under Facebook's incitement rules, which also prevent certain types of disinformation posts that could cause offline harm. The Oversight Board, basing its decision in part on international human rights law, ordered the post to be reinstated because Facebook had not demonstrated the post caused a risk of "*imminent* harm."[64] And indeed, were a government to have a blanket ban on borderline posts such as this, it might run afoul of applicable international law and certainly the First Amendment.

But Facebook ought to be able to strike the balance in a different way. It is being blamed for being a cauldron of COVID misinformation, leading people to doubt vaccine effectiveness, to experiment with dangerous "cures," and to believe a host of conspiracy theories. If Facebook decided to filter out all talk of hydroxychloroquine, given the risk that some people might die from prophylactic use of an unapproved drug, it would certainly prevent some valuable medical information from reaching its audience (such as warnings not to take the drug). But again, Facebook is not the only place on the internet where one might find information or misinformation on COVID. Its cost-benefit analysis on the harms of overcensoring through automated filtering versus potential lives saved through avoidance of COVID disinformation should be different than that of a government.

Evelyn douek is one of the few scholars who have recognized this point.[65] She argues, at a length unavailable here, that the international law rule of "proportionality" in speech regulation needs to be adapted to "probability" determinations that platforms can use in their automated and human moderation systems, especially including the algorithms that determine the reach of content. As a result, platforms' overbroad rules more easily adapted to automated global enforcement require, at least, a rethinking of what might be a new international human rights norm for platform content moderation.

Platform Rights and Platform Obligations

How, then, should we think about the rules that platforms should impose on users and that governments should impose on platforms? As should be clear from the discussion presented here, both of these decisions need to be informed by the unique features that distinguish large platforms from governments: namely, the global environment in which they operate, the speed with which they need to make decisions, the fact that prior restraints (usually through automated filtering at the source) represent the only effective way of preventing many online harms, and the special character of algorithms as a source for organizing and prioritizing speech.

As a threshold matter, we should cast to the side the notion of strong First Amendment rights for large platforms. They may not be common carriers, as argued above, but so too they are not like a newspaper or private club. These corporations' First Amendment rights (in the *Citizens United* sense) should be limited by important state interests in preserving democracy, preventing online and offline harms, and ensuring a competitive and healthy online marketplace of ideas. The First Amendment should be read as preventing many broad regulations of speech on the internet, but when it comes to regulating the large platforms themselves, the government should have much greater latitude. Many regulations of platforms might violate users' First Amendment rights (e.g., forcing Facebook to take down all insults, would, in fact, make them the state's agent in doing so),[66] but the platforms' own rights should not prevent these regulations.

Even when constitutional, such regulations might not be wise, of course. For reasons explained above, an "all comers" policy based on common carrier theories would cause more problems than it would solve. However, more modest antidiscrimination approaches could succeed. For example, whereas small platforms and websites have the right to discriminate on the basis of race or gender, it seems reasonable and constitutional to prevent the large platforms from doing so. The same could be said with respect to declared candidates and parties: The law could prevent the large platforms from using their power to exclude one or another candidate. Of course, that does not mean they can't disproportionately take down the accounts or content of Republicans or Democrats if, for example, one party tends to engage in disinformation and hate speech. Nor does it mean that each user must receive an equal amount of Democratic and Republican content; the platform could still bias the delivery of content based on user behavior and search history. It just means that platforms of a given size cannot use their monopoly position to intentionally favor or disfavor a candidate or party without some other community standard enforcement reason for doing so.

Moreover, a whole suite of regulations of advertising on large online platforms should be seen as presenting no constitutional problems. Requiring total transparency of purchasers and advertising contracts, bans or limits on microtargeting, and unique regulations for political advertising are examples of reasonable regulations. Indeed, a host of transparency rules, which might be constitutionally problematic if applied by government to the entire internet, should be seen as constitutionally unproblematic when applied to the large platforms. This would include compelled access to aggregated platform data by outside researchers, as well as periodic audits of algorithms to avoid (or at least be transparent about) bias and to ensure safety. I have authored one such piece of legislation, the Platform Transparency and Accountability Act,[67] which would empower the Federal Trade Commission to compel platforms to develop secure, privacy-protected pathways for independent research on platform-controlled data. The goal of such legislation is not to provide a luxury good to academics granted access but, rather, to alter platforms' behavior by removing their ability to operate in secret and to provide reliable information to policymakers for further legislative initiatives.

When it comes to moderation of content such as disinformation and hate speech, there is not much the government can do without running afoul of the First Amendment rights of users. A White House Office of Information Integrity is something that should send shivers down bipartisan spines. If the government is going to regulate harmful but legal content, it will need to do so indirectly. For example, it might require large platforms to comply with their own stated policies regarding content moderation. This suggestion may be a bit more controversial, given that many of their policies would violate the First Amendment if applied by government. Indeed, this principle was in part the basis for President Trump's controversial executive order following Twitter's actions against his posts.[68] As loony as the executive order was in other respects, its suggestion that repeated misapplication of explicit content moderation rules might constitute a kind of fraud presents a creative potential middle path for enforcement.[69] It might lead the platforms to promise less in terms of content moderation or perhaps all their rules would become mere suggestions. But given how difficult (and as argued above, wrongheaded) legislation on content moderation might be, this attempt at "co-regulation" (to borrow from the European approach) may hold some potential for striking the right balance.

When it comes to the platforms themselves, we should not expect or desire each platform to resolve controversial issues in the same way (in terms of both process and result). Indeed, as evelyn douek warns, there are good reasons to fear "content cartels."[70] Moreover, the "affordances" of the platforms, let alone the resources they have to dedicate to content moderation, differ considerably. Video platforms like YouTube and TikTok face different challenges than a social media company like Facebook, which, given its size and wealth, has very different capabilities than does Twitter.

The decisions of the Facebook Oversight Board are important because they are the first official body with some distance from the platform itself (but with real-world impact) to grapple openly with the challenges of content moderation. We need more institutions like it, whether as a product of industry associations, corporate-government partnerships, or single-firm initiatives. The more transparency we have around resolution of the hard cases of content moderation, the more we might be able to lurch toward something resembling a common law in this space.

As described above, I think the common law here looks different from what we might see from constitutional courts. It will still balance individual rights of expression against public interests in preventing harm, but it will also include legitimate corporate interests tailored to the functioning of the product. These are not the raw financial interests of the company (although one can legitimately debate whether those are in scope as well) but, rather, the kind of speech environment that the platform seeks to cultivate. Are children using the platform? Are there closed groups, or are all posts viewable by all users? How much power is given to individual users to determine the content they see? (Indeed, this last question may hint toward a future of distributed content moderation, in which a market for different rules develops and users can opt into different community standards or different algorithmic prioritization rules. This, along with larger moves toward interoperability between platforms, might change the way we think of the impact of platform rules on speech.)

This common law, however, must be tailored to the unique challenges of rapid, automated content moderation on a global scale. Those features of platform speech regulation, as compared to that done by governments, necessarily translate into error rates at levels that constitutional courts would not and should not tolerate. This warranted tolerance for higher error rates comes, in part, from the fact that when a government (unlike a platform) bans speech or speakers, it often (though not always) removes their ability to utter those words anywhere in public or at least from the *truly* public square.[71] Decisions by firms and outside oversight bodies, let alone chatter from the pundit class, need to account for the place of a platform's speech rules in the larger context of online speech. The reason why large platforms' own speech rights should be given short shrift (as I argued above) is the enormous (arguably oligopolistic) power a few firms wield over the information ecosystem. Nevertheless, when certain content disappears from Twitter, for example, it does not disappear from the internet. Even when speech is taken down, it can almost always be placed elsewhere for users to view.

Of course, loss of an audience on one of the major platforms can be devastating for speakers. Creators on YouTube can spend years cultivating a following,

only to see it vanish overnight if they get deplatformed or otherwise demoted. Similarly, President Trump's efforts to re-create the audience he enjoyed on the major platforms fell flat. Costs measured by reach, instead of speech, are still significant costs, and may mean the loss of livelihood for creators and publications that have become dependent on a platform. But for most enforcements of community standards, as in the Oversight Board cases described above, the platform's decision does not prevent the speakers from getting their message out; it merely prevents them from taking advantage of the platform algorithm that amplifies it.

In the end, though, rising platform power may present the greatest challenge to the speech marketplace. That problem needs to be addressed head-on through antitrust and competition regulation. Indeed, the debate over content moderation cannot be divorced from the antitrust debate: If there were twenty small Facebooks, we might be less concerned over the community standards enforcement of any given one. For now, though, we should aim for a more transparent and experimentalist enforcement regime from the large platform from which a common law of content moderation might develop.

Conclusion

The summer of 2021 may be remembered as a turning point when it comes to regulation of social media. The Facebook whistleblower Frances Haugen captivated the public and Congress, first with a series of disclosures of Facebook's internal documents published in the *Wall Street Journal* and then with testimony before a Senate committee. She alleged that Facebook put profit before safety, and came with examples drawn from internal studies and documents relating to everything from Instagram's effect on teen girls' health to special content moderation appeals processes for influencers to negligence in policing offline harm in the global South. Haugen's testimony has lit a spark that may finally lead to comprehensive legislation to deal with the assorted problems afflicting the large platforms. At a time of historic partisan polarization, it appears that mutual hatred of "big tech" may be one of the few things that unites Democratic and Republican lawmakers.

Then again, the parties remain far apart when it comes to their characterization of "the problem" and therefore the desirable solutions. Democrats tend to want the platforms to take down more content, especially disinformation, hate speech, and incitement, and Republicans worry that aggressive content moderation policies lead to censorship of conservatives. Finding a bipartisan, constitutionally viable regulation of harmful content will not be easy, except maybe in a few narrow contexts such as child endangerment. Rather, we should expect

legislation on privacy, competition, taxation, advertising, and transparency to have a greater likelihood of passage.

In contrast to the typically sclerotic US policymaking process, Europe may be a first mover when it comes to tech policy. As scholars have shown in other contexts of "the Brussels Effect,"[72] European regulators often set rules with global extraterritorial application. We have seen this in the tech context with the General Data Protection Regulation (GDPR), Europe's privacy law, which companies like Facebook have said they will apply worldwide. The same might happen with content moderation as Europe begins to draft the specifics for enforcement of the Digital Services Act, which contemplates broad regulation of American tech companies. Without speech protections as robust as those in the US First Amendment tradition, Europe may be able to impose more significant content moderation standards. In turn, the US tech companies may (as with GDPR) decide that it is easier to apply such rules worldwide than have different standards by region.

For now, we need a better understanding of what is happening on these platforms and how platform policies are shaping the information ecosystem.[73] We should not need to wait for whistleblowers to blow their whistles in order to get glimpses of the insights platform researchers have culled from company data. The flip side of freedom of expression is a right to information. The public has a right to be informed about what is happening on these large internet platforms. However one might characterize these new, powerful entities for purposes of constitutional law, these platforms have lost their right to secrecy. Congress can and should require that they open themselves up to outside scrutiny.