OCCASIONAL PAPERS



Amplification and Its Discontents

By Daphne Keller

KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY

The Knight Institute's Occasional Papers series aims to bring to a broad audience thoughtful, provocative work from scholars and experts who usually write for more specialized audiences. The open-ended series features papers that address urgent questions at the intersection of speech, privacy, and technology.

The series is conceptualized by Knight Institute staff, including Jameel Jaffer, Executive Director; Katy Glenn Bass, Research Director; Alex Abdo, Litigation Director; and Larry Siems, Chief of Staff. The series is edited by Glenn Bass with support from Lorraine Kenny, Communications Director; Adam Glenn, Writer/Editor; Chand Rajendra-Nicolucci, Research Fellow; and Madeline Wood, Communications and Research Coordinator.

The full series is available at knightcolumbia.org/research/

I. INTRODUCTION

THERE IS A POPULAR LINE of reasoning in platform regulation discussions today that says, basically, "Platforms aren't responsible for what their users say, but they are responsible for what the platforms themselves choose to amplify." This provides a seemingly simple hook for regulating algorithmic amplification—the results for searches on a search engine like Google or within a platform like Wikipedia; the sequence of posts in the newsfeed on a platform like Twitter or Facebook; or the recommended items on a platform like YouTube or Eventbrite. There's some utility to that framing. In particular it is useful for people who work for platforms building product features or refining algorithms.

For lawyers or policymakers trying to set rules for disinformation, hate speech, and other harmful or illegal content online, though, focusing on amplification won't make life any easier. It may increase, rather than decrease, the number of problems to be solved before arriving at well-crafted regulation. Models for regulating amplification have a great deal in common with the more familiar models from intermediary liability law, which defines platforms' responsibility for content posted by users. As with

ordinary intermediary liability laws, the biggest questions may be practical: Who defines the rules for online speech, who enforces them, what incentives do they have, and what outcomes should we expect as a result? And as with those laws, some of the most important considerations—and, ultimately, limits on Congress's power—come from the First Amendment. Some versions of amplification law would be flatly unconstitutional in the U.S., and face serious hurdles based on human or fundamental rights law in other countries. Others might have a narrow path to constitutionality, but would require a lot more work than anyone has put into them so far. Perhaps after doing that work, we will arrive at wise and nuanced laws regulating amplification. For now, I am largely a skeptic.

In this essay, I will lay out why "regulating amplification" to restrict distribution of harmful or illegal content is hard. My goal in doing so is to keep smart people from wasting their time devising bad laws, and speed the day when we can figure out good ones. I will draw in part on novel regulatory models that are more developed in Europe. My analysis, though, will primarily use U.S. First Amendment law. I will conclude that many models for regulating amplification face serious constitutional hurdles, but that a few—grounded in content-neutral goals, including privacy or competition—may offer paths forward.

This assessment draws in part on my own experiences with both manual and algorithmic management of content at Google, where I was associate general counsel until 2015. In that capacity, I advised on compliance with many content-restriction laws around the world, and spent a great deal of time with the engineers who build the company's ranking algorithms. I will do my best to flag where my own policy preferences—which both led me to that job and were, presumably, shaped by it—influence the analysis set forth here.

A. Amplification and harm

Amplification features can do both harm and good. At the beneficial end of the spectrum, they help us find information on the web or within individual sites. Ranking on platforms like Twitter and recommendations on platforms like Etsy help users discover new content, goods, artists, activities, and ideas.

But major platforms' amplification features have also caused or

contributed to real damage in the world. At a societal level, they have spread misleading political material, to the detriment of democratic governance.[1] At an individual level, they may lead dieters to content promoting anorexia, or viewers of Trump rallies to videos denying the Holocaust.[2] Facebook's friend and group recommendation algorithms are said to have brought together violent right-wing extremists, one of whom ultimately shot and killed two people in Kenosha, Wisconsin.[3]

This essay will examine potential legal models for harnessing the benefits of amplification, while reducing the attendant harms. My focus here is on the problem of harmful or illegal content posted by internet users, and the risk that this content causes still more harm when it is amplified by platforms. That means I will not be examining some important, but conceptually distinct, concerns that amplification contributes to other legal or societal problems, beyond the spread of harmful or illegal content. Important issues that are out of scope include:

- Ranking or amplification that is itself *discriminatory*. This might include withholding otherwise lawful offers of housing, employment, or credit based on a user's race, as Facebook is alleged to have done.[4] It might also include showing all white men in a search for professors.[5] The problem in those cases is not that housing ads or pictures of white men in tweed are inherently bad. It's that platforms introduce harm distinct from that content through their ranking or targeting. These and other instances of algorithmic bias are huge concerns, but out of scope for this essay.
- Ranking or amplification that is *anti-competitive*, as the European Commission concluded was the case with Google's ranking for its own shopping service in web search results.[6] These cases, too, involve showing otherwise-innocuous content in the wrong place.
- Ranking or amplification that is harmful for *privacy or data protection* reasons, because of the way it leverages user data. As I will discuss in Section IV, laws grounded in privacy may be valuable in responding to the amplification of harmful content. But I will not examine purely privacy-based harms or the laws that might remedy them.

This essay will define "amplification" to encompass various platform features, like recommended videos on YouTube or the ranked newsfeed on Facebook, that increase people's exposure to certain content beyond that created by the platform's basic hosting or transmission features.[7] I will use the term "demote" to cover any form of deamplification, including decreasing content's algorithmic ranking or excluding it from features like recommendations. My focus will be on "organic" user-generated content (not the content of advertisements), and on consumer-facing platforms like Facebook or YouTube (not infrastructure providers like CloudFlare or Amazon Web Services).

This definition conflates a few categories that could, in other contexts or a longer essay, be usefully distinguished. For one thing, it encompasses purely user-initiated virality, like widespread sharing of electoral disinformation on platforms like WhatsApp, as well as the additional algorithmic boost platforms might provide. It also includes both "pull" models like the search results a user requests from Google and "push" models like YouTube video recommendations.[8] Finally, it includes both actions platforms take in response to specific content (like demoting news items identified as false by fact checkers) and global algorithmic changes (like Google's 2017 shift to reduce ranking of content including "hoaxes and unsupported conspiracy theories").[9]

The concept of internet amplification may inevitably be fuzzy at the edges. Almost any act that spreads or draws attention to particular information could be characterized as amplification. Thinking about the exact meaning too hard can send a person down mental rabbit holes. You could say that putting content anywhere on the internet amplifies it to a wider audience than it might find in print. Or that being on a high-traffic service like Facebook amplifies content, bringing it more attention than it would get on most websites. Or that content that is prominently positioned in Facebook's user interface (like the middle of the screen) is amplified compared to other parts of the page (like the bottom edge). The difficulty in drawing crisp boundaries can make the distinction between "publishing" and "amplifying" seem indeterminate or artificial. The well-known difficulty of defining authentic or "neutral" platform behavior finds its mirror image in the difficulty of defining artificial or "amplifying" behavior.[10] I will explore that issue more deeply in

the final section of this paper. Overall, though, I believe "amplification" has a rough conventional meaning assumed in today's discussions, and that is the one I will try to use.

B. Three models for regulating amplification

Lawmakers around the world have proposed regulating platforms' amplification of online content. A proposal backed by the Trump administration, for example, would have stripped platforms of immunity for user content that they algorithmically "promote[d.]"[11] An EU Parliament draft report discussed restricting "the amplification of content that is attention-seeking or sensationalist in nature."[12] Other proposals seek instead to harness amplification tools. One draft German law would require promoting diverse voices.[13] The European Commission has supported the perhaps-conflicting goal of promoting authoritative ones.[14] A Facebook experiment along these lines demoted posts rated as false by third-party fact-checkers, leading to an average 80 percent reduction of viewership.[15] Interventions of this sort can allow social media platforms and policymakers to choose more nuanced responses than the binary take down/leave up choices recognized under most laws today.

These proposals build on recognition that, as my Stanford colleague Renée DiResta put it, "Free speech is not the same as free reach."[16] Platforms that host user speech can opt not to give it additional reach via features like recommendations. Private companies have no obligation to host their users' speech, *or* to provide it with additional reach via amplification. They can opt to permit one without the other. Lawmakers do not have that same freedom. With narrow exceptions, the government can't restrict speech *or* reach. That makes regulation challenging.

The main sections of this essay will analyze three general approaches to this challenge. I look forward to interesting conversations with fellow wonks about these, especially those who may disagree with my assessments. Policymakers and people in a hurry, though, may want to skip to Section IV. The "privacy" and "competition" options discussed there are the approaches that I think are most promising, with the fewest First Amendment problems.

To preview, here are the three kinds of models.

1. ILLEGAL SPEECH MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING ILLEGAL SPEECH (SECTION II)

Policymakers could increase platforms' liability for amplifying illegal content posted by users. I will discuss these "illegal speech models" in Section II. For example, platforms could lose immunities currently available under laws like Communications Decency Act 230 ("CDA 230") if they amplify defamatory posts.[17] Supporters might reason that such a law should face lower constitutional barriers, since it only affects amplification features—it does not require platforms to remove speech from the platform entirely. As I will explain, though, a long line of U.S. Supreme Court precedent goes against this reasoning. Laws that reduce visibility of speech face the same strict scrutiny under the First Amendment as laws that ban it outright. New liability for amplification would thus raise the same constitutional question as any other intermediary liability law: how to define platforms' obligations in a way that minimizes resulting collateral damage to users' lawful speech. If we believe this problem is solved, and that platforms can identify illegal speech reliably enough, the appropriate policy response would seemingly be to require them to delete it—not just stop amplifying it. If we do not believe the problem is solved, then laws regulating amplification will raise the same questions as the more familiar laws that regulate content hosting or transmission.

2. HARMFUL SPEECH MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING CURRENTLY LAWFUL BUT HARMFUL SPEECH (SECTION III)

A major premise for proposals to regulate amplification is that certain content becomes more harmful when spread to a broader audience online. Disinformation, for example, poses a greater threat to elections if it reaches more voters. Arguably, this increase in harm should shift the constitutional calculus—allowing lawmakers to restrict otherwise-lawful material if it appears in features like YouTube recommendations. I'll call laws like this "harmful speech models" and discuss them in Section III. European laws provide some precedent for this approach and lessons about its difficulties. One is that defining a new margin of restricted content beyond that already known to legal experts is a major undertaking. Establishing an unprecedented set

of new speech laws and adequately resolving interpretive disputes would require administrative capacity well beyond that of any current U.S. regulator (or, I expect, any European one). I will also discuss U.S. precedent, which generally grants little leeway for laws of this sort. Arguably, though, a path forward could be crafted based on the rules that the Supreme Court has upheld for older media like broadcast or cable.

3. CONTENT-NEUTRAL MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING ANY SPEECH (SECTION IV)

Finally, I will discuss approaches that avoid at least some of the major constitutional problems created by the illegal or harmful speech models. Lawmakers could try to define rules that avoid preferencing one message over another—that are, in legal parlance, content-neutral. In Section IV, I will discuss a few of these "content-neutral models." One, eliminating amplification outright, strikes me as extreme and particularly vulnerable to First Amendment objections. Others might be more viable. A "circuit-breaker" approach that uses numerical limits to dampen the viral spread of any content, for example, might pass constitutional muster and be effective in addressing certain amplification-based harms. Other approaches grounded in privacy, competition, or, potentially, consumer protection law strike me as most promising. They might address some of today's concerns while avoiding many of the constitutional difficulties otherwise involved in regulating amplification of online speech. They might also cause the most pain for today's largest incumbent internet platforms. When Facebook says it welcomes regulation, this is not the kind it is talking about.[18] So, while the constitutional barriers to this kind of regulation may be lowest, the litigation and lobbying effort against it might be highest. Major platforms, which already invest heavily in content moderation, may well prefer the expense—and correlating competitive advantage—of the illegal or harmful speech models.

II. ILLEGAL SPEECH MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING ILLEGAL SPEECH

In "illegal speech" models for regulation, lawmakers would restrict amplification of particular unlawful user content, or remove otherwise available statutory immunities. At least one proposed U.S. law has taken this approach, eliminating immunity under CDA 230 for terrorism or civil rights lawsuits about amplified content.[19] Similar ideas seem likely to follow.

As I will explain in this section, there are several constitutional or fundamental rights-based issues with such laws.[20] Some—the more interesting and important ones, I think—involve *users'* rights to share lawful speech online. For users' rights, two separate strands of First Amendment jurisprudence matter. First, laws restricting distribution of speech face the same strict First Amendment scrutiny as laws banning it. That means that laws requiring platforms to demote content have the same problems as laws requiring them to remove it. Second, a law penalizing amplification would have the same problem as more familiar laws that penalize platforms for carrying content at all: They encourage platforms to over-enforce and suppress lawful speech. A law that pushes platforms too far in that direction can violate the First Amendment—regardless of whether the result is that legal speech gets demoted, or deleted altogether.

A second set of issues involves the First Amendment rights of *platforms*. Platforms' own arguments against illegal speech models for regulating amplification are relatively weak. But those arguments will recur—as will the arguments based on users' rights—in different and sometimes stronger forms under the other two legal models discussed later in this essay.

A. User rights issues

1. THE *PLAYBOY* RULE: BURDENS AND BANS ON INTERNET USER SPEECH GET THE SAME SCRUTINY

A law telling platforms to demote or promote particular kinds of content, or holding them liable for failure to do so, would be a law regulating platform

users' speech. Framing that law as a mere "restraint on amplification" instead of an outright prohibition would not change the applicable constitutional scrutiny. The Supreme Court has spoken firmly and repeatedly to this point. In *U.S. v. Playboy*, a case striking down requirements for cable operators to limit access to the plaintiff's pornographic content, it wrote:

> It is of no moment that the statute does not impose a complete prohibition. The distinction between laws burdening and laws banning speech is but a matter of degree. The Government's content-based burdens must satisfy the same rigorous scrutiny as its content-based bans.[21]

The Court has applied similar reasoning in striking down other laws that burdened speech without banning it. In the offline world, it rejected Cincinnati's attempt to exclude print advertising flyers from distribution in favorably positioned newsracks.[22] It also struck down New York's "Son of Sam" law, which effectively demonetized criminals' memoirs by diverting sales revenue to victims.[23] Notably, some of these older laws—like many proposals about internet amplification today—directly regulated intermediary companies, but effectively restricted speech by third parties. The law in *Playboy* facially regulated cable companies, for example, but functionally burdened speech by their programming providers. Programmers successfully challenged the law as a violation of *their* First Amendment rights—not the cable companies' rights. In the same way, regulations that facially regulate platforms today may functionally burden constitutional rights of their users.

Lower courts have applied the same principles to the internet, and even rejected government efforts to permit speech on one part of an online service while excluding it from other parts of the same service. The U.S. Circuit Court of Appeals for the D.C. Circuit, for example, held that the Federal Election Commission could not constitutionally restrict more speech in the title of a webpage than it did in the same page's less-conspicuous body text.[24] As that court noted, "talk[ing] about a candidate in the body of a website is of no use if no one reaches the website."[25] The Second Circuit, similarly, rejected the Trump administration's claim that because Twitter users could still use less visible parts of the service, the president could permissibly exclude them from the higher-profile discussions among followers of his account.[26] The

Supreme Court's ruling in *Packingham v. North Carolina* applied similar logic in striking down a restriction on sex offenders' access to social media sites. It rejected—even under the looser "intermediate" level of First Amendment scrutiny—the lower court's conclusion that restricting speech on popular social media sites is OK, because people can still speak on other websites.[27]

Of course, even strict First Amendment scrutiny is not *always* the kiss of death for speech regulations. A sufficiently narrowly tailored law could survive judicial review, as I'll discuss later in this section. First, however, I will describe the main speech-related issue that would arise for amplification laws, as it does for nearly all intermediary liability laws: platforms' incentives to over-enforce and suppress lawful speech.

2. THE *SMITH* RULE: LEGALLY MOTIVATED OVER-ENFORCEMENT BY PLATFORMS THREATENS USERS' SPEECH RIGHTS

Laws requiring platforms to demote certain content would have the same practical problems as laws requiring them to delete it: Platforms are bad at identifying which user speech violates the law and not well-motivated to try harder. This is a familiar problem for anyone who follows the inconsistent results of platform content moderation—a generally simpler task, in which platforms apply content rules they themselves devised. When required to interpret more nuanced legal rules under threat of liability, platforms' performance is also poor. They tend, predictably, to protect themselves by erring on the side of over-enforcement. The resulting suppression of lawful speech has constitutional significance in the U.S., and raises issues of human and fundamental rights in other legal systems. This is a topic I've written about extensively elsewhere, so I will address it only briefly here.[28]

For legal systems outside the CDA's blanket rules, a common approach is to grant platforms immunity unless they know about particular illegal content. This is how copyright law works under the U.S. Digital Millennium Copyright Act (DMCA), and how liability for unlawful user content of any kind works under the EU's e-Commerce Directive. Empirical research has documented considerable over-enforcement by platforms taking down legal speech under such laws in order to avoid expense or legal risk for themselves.[29] The problem is compounded by "heckler's veto" attempts by notifiers who submit false legal allegations to platforms. The government of

Ecuador, for example, notoriously used bogus copyright notices to get major U.S. platforms to remove critical news reporting and videos of police brutality.[30] Other abusers have used the same trick to squelch reporting about fraud and professional wrongdoing in the U.S.[31] We should expect problems of this sort to be no less prevalent under laws requiring platforms to demote particular content. And, following *Playboy* and other cases discussed above, the resulting burden on user speech would be as constitutionally significant as it would be if platforms deleted user speech, instead of demoting it.

Intermediary liability laws can also shape speech rights in more subtle ways. As the European Court of Human Rights has noted, poorly crafted laws can deter investment in building open platforms in the first place.[32] They can also nudge existing platforms to simply prohibit a broader swath of user behavior using their terms of service. That approach simplifies operations, reducing both costly legal review operations and litigation risk. That kind of platform response to legal pressure, which is already hard to detect, would be effectively invisible if the sole public evidence were algorithmic demotion of speech potentially disfavored by governments around the world.

Intermediary liability laws almost always facially bar only unlawful content. But they can still violate the First Amendment and its international analogs if they go too far in motivating intermediaries to over-enforce. The Supreme Court made this clear in a case about bookstores, *Smith v. California*. There, the Court rejected strict obscenity liability, noting that the bookseller's resulting timidity" from such unbounded liability can lead it to "restrict the public's access to forms of the printed word which the State could not constitutionally suppress directly." The resulting "censorship affecting the whole public" would be "hardly less virulent for being privately administered."[33]

One of the most forceful judicial statements of this reasoning comes from an otherwise largely forgotten Eighth Circuit ruling. In *Midwest Video v. FCC*, that court rejected an FCC regulation requiring cable operators to restrict unlawful content from programmers. It pointed out that

> [i]n so mandating, the Commission appears to have created a corps of involuntary government surrogates, but without providing the procedural safeguards respecting "prior restraint" required of the government. ...
> Thus the Commission made the cable operator both judge and jury, and

> subjected the cable user's First Amendment rights to decision by an unqualified private citizen, whose personal interest in satisfying the Commission enlists him on the "safe" side—the side of suppression.[34]

The same concern arises today about laws that would make platforms the "judge and jury" for legal decisions about online speech they host or amplify. What can be done to minimize the resulting risk of over-enforcement? Or, in the language of First Amendment strict scrutiny: How might lawmakers more "narrowly tailor" the law to improve the fit between governments' goals and the means chosen to achieve them? Major models in human rights literature and existing U.S. law use procedural rules to improve platforms' ability to correctly identify which content violates the law.[35] The U.S. DMCA and Europe's pending Digital Services Act, for example, use choreographed notice-and-takedown processes. Procedure-based laws of this sort can require platforms to notify users when taking down allegedly illegal speech, and give users opportunities to appeal those decisions, for example. (Informing affected users would be all the more important if laws caused platforms to demote content, since users might not notice that change.[36]) Laws can also penalize bad faith "heckler's veto" allegations, and give users recourse to courts. Experience with the DMCA tells us that such rules aren't enough to ensure good outcomes. But, like civil procedure in courts, they can provide a record and set incentives to make good outcomes more likely.

Other legal systems mitigate over-enforcement problems by involving courts or regulators.[37] In some countries, legislation and case law provides that platforms should never be the legal decision-makers for complex questions of law or fact. Instead, to protect users' speech rights, courts or public authorities must decide which speech is illegal, providing due process to the speaker. Then platforms can be served with court orders, and lose immunity if they fail to take down the content. Brazil's Marco Civil, for example, requires a court to review alleged defamation, but makes platforms responsible for assessing more recognizable and severely harmful content like child sexual abuse images. Any one of these approaches, borrowed from ordinary intermediary liability laws, might provide a step toward better tailoring amplification laws.

3. PUTTING *PLAYBOY* AND *SMITH* TOGETHER

The *Playboy* line of cases and the *Smith* line of cases together support this conclusion: A content liability law that encourages over-enforcement by platforms faces the same First Amendment scrutiny, regardless of whether the law regulates platforms' amplification or basic hosting. That does not, in principle, preclude different legal treatment of amplification—if it can survive strict scrutiny. The relative rigidity of U.S. First Amendment analysis, though, makes it hard to arrive at policy resolutions that may be more acceptable in some other parts of the world.

A law restricting amplification might, in some legal systems, be defended based on the balance of harms. When widespread replication of already-illegal content creates more danger, proponents might argue, the state's interest in restricting amplification becomes even more compelling than the interest that justified prohibiting that material in the first place. Averting this greater harm might justify accepting some additional margin of platform over-enforcement against lawful speech. One could also reason that collateral damage to lawful speech is less significant when platforms merely exclude that speech from recommendations or ranked newsfeeds, but continue to host it. Or even that lawful speech that can be mistaken for illegal content was probably not very valuable, and thus weighs less heavily in the balance.

For better or for worse, this kind of analysis is hard to reconcile with black-letter First Amendment doctrine.[38] U.S. case law tells us that, regardless of whether lawmakers ban speech or merely burden it, requirements to narrowly tailor the rules to minimize incidental damage to lawful speech remain the same. The amplification cases discussed above all failed in this tailoring requirement. The cable pornography-blocking law in *Playboy*, for example, was rejected because Congress could have chosen alternate approaches that would have avoided making cable companies block lawful pornography transmissions.[39] As the Supreme Court emphasized, this is a technology-specific question. If new technologies permit more accurate targeting of unlawful speech, Congress should use them.[40]

Do platforms' amplification algorithms constitute such a technology—do they render platforms more capable of recognizing unlawful speech and thus justify assigning them greater responsibility for doing so? Such

a conclusion would be, in my opinion, very premature. Platforms may talk a big game about the AI or machine learning behind their algorithms. But neither experience nor research suggests that algorithms can reliably distinguish legal from illegal content, outside of very limited cases. Amplification algorithms can succeed by being scattershot and imperfect.[41] That's why I—an attorney—see so many ads suggesting that I enroll in law school. That shotgun approach is good enough to make platforms money. But it is ill-suited to the more delicate task of legal judgment.

When platforms do try to identify unlawful material using algorithms, they regularly fail. Software can be good at discerning broad patterns, and it is increasingly good at identifying both duplicates and near-duplicates of images or videos. But it is bad at discerning the meaning of human communications. Other than child sexual abuse material—which is both uniquely harmful and never legal in any context—nearly any image, video, or text that violates the law in one situation can be legal in another. Algorithms designed to find terrorist material, for example, can't tell the difference between ISIS propaganda and news reporting. Algorithmic failures have been blamed for serious mistakes including YouTube's deletion of videos documenting war crimes in Syria.[42] An increasing body of evidence also suggests that errors of this sort are not evenly distributed, but disproportionately penalize certain people—including, for example, speakers of African American vernacular English.[43] U.N. human rights officials and civil society organizations have, for these reasons, strongly opposed proposals to make platforms use algorithms to identify illegal content—even when coupled with human review of machines' output.[44]

The problems with tying liability to ranking algorithms get worse if we imagine a law that, instead of targeting amplification of specific prohibited content such as posts by terrorist organizations, instead eliminates immunity for amplifying *any* unlawful content. Since the algorithms behind features like newsfeeds or recommendations rank *every* item, this would effectively expose platforms to liability under any law for anything users post. A rational platform would presumably respond by either eliminating amplification features entirely, or excluding broad categories of content—much as Craigslist eliminated its personal ads and many other platforms cut off service to sex workers in response to a 2018 law eliminating platform immunity for

claims involving prostitution or sex trafficking.[45] For speech or users on platforms viewed primarily via ranked feeds, these changes would amount to banishment.

A law broadly penalizing and discouraging amplification would also be odd as a legal matter. Americans generally *want* platforms to engage in content moderation.[46] Congress enacted CDA 230 in order to encourage moderation, including takedown of "objectionable" but lawful content. Even legal systems like the EU's, which nominally require platforms to be "passive" or "neutral" in order to be immunized, have awarded immunity to algorithmically ranked features like Google's AdWords—and have in recent years moved emphatically to, like the U.S., encourage platforms to moderate content.[47] It seems perverse to encourage moderation in the form of binary take down/leave up decisions, but discourage more nuanced responses involving ranking or targeting.

The limitations of today's technology may not last forever. Perhaps a day will come when algorithms really do give platforms enough information that they should reasonably be expected to parse and assume responsibility for user speech. When that day comes, though, presumably we should require platforms to *remove* the illegal content they learn about—not just stop amplifying it. In the meantime, the algorithms responsible for amplification do not meaningfully change the practical and constitutional options for platform regulation.

4. PLATFORM RIGHTS ISSUES

A second set of constitutional issues comes not from users' First Amendment rights but from platforms' own. These arguments are not terribly compelling, I think, as responses to the illegal speech model discussed in this section. But they become more so under the harmful speech and content-neutral models discussed later in this essay.

Platforms that use algorithms to rank user content effectively set editorial policy and "speak" through ranking decisions. The message conveyed can be pretty boring: Platforms say things like "I predict that you'll like this" or "I think this is what you're looking for." That's enough that lower courts have recognized First Amendment protection for platforms' ranking choices.[48] I think they are likely to continue doing so, barring a Supreme

Court-level change in jurisprudence, because the Court has set a low bar in defining First Amendment rights of entities that aggregate third-party speech. It has, for example, recognized protectable First Amendment interests in cable companies' selection of programming, and even in a parade operator's purely negative choice to exclude particular participants.[49] The difference between user speech and platform speech is analogous to the difference between an essay and the anthology that contains it—each of which is deemed a distinct creative work under U.S. copyright law, with the latter receiving its own protection based on the anthologist's selection and arrangement of third-party speech.[50]

It could be argued that a law incentivizing platforms to over-enforce and remove lawful content from recommendations or newsfeeds would affect platforms' rights in much the same way as users' rights. Every time a platform removes lawful user speech out of caution and fear of liability, the argument would go, it is also self-censoring and forgoing its own right to recommend or convey that lawful content. It's hard to muster much sympathy for the platform's problem, though. Compared to users, the platform has far more agency. It makes its own mistakes, while users are subject to decisions outside their control. Wealthier platforms could, up to a point, reduce mistakes by hiring more lawyers to review content. So the arguments based on platform speech rights seem, under the illegal speech model, not as important or strong as those based on user speech rights. As I said, though–keep an eye on this argument. It will come back later.

III. HARMFUL SPEECH MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING CURRENTLY LAWFUL BUT HARMFUL SPEECH

THE NEXT POTENTIAL MODEL I'll discuss goes a step further: restricting amplification of speech that is currently legal. The rationale for this approach stems from the main critique of amplification—that certain speech becomes more dangerous when it spreads rapidly on the internet. If we take this idea seriously, then restricting previously legal lawful speech, like legal hate speech or medical misinformation, might

make sense. That approach would raise all the same issues about platform over-enforcement as the illegal speech model, discussed in Section II. I won't repeat those here. But it would also raise new issues about defining which previously legal speech must now be restricted.

Perhaps unsurprisingly, proposals to restrict online propagation of harmful but lawful speech have had the most traction outside the U.S. As I will discuss in this section, the Right to Be Forgotten in the EU provides an illustration. The U.K. is also pretty far down the road to restricting lawful but "harmful" material online, under rules that will be overseen by the equivalent of the Federal Communications Commission.[51] A similar approach in the U.S. seems unlikely, but not impossible, given the second topic I will discuss below: our own history of communications regulation, including Supreme Court First Amendment case law.

A. European models

The first major EU legal development tackling amplification-based online harms was the 2014 *Google Spain* case, which defined what is now known as the Right to Be Forgotten.[52] The EU's Court of Justice in that case determined that certain personal information about individuals can be legal for webmasters to publish online, but not for search engines to distribute more widely via search results.[53] Search engines had to "delist" that information from the results for certain queries upon request. The ruling effectively added friction to the spread of online information, slowing its distribution via high profile intermediaries without banning it from being published or hosted elsewhere. Thus, while the doctrinal basis was very different, both the consequences and the Court's analysis about harm from "ubiquitous" content functionally resembled recent proposals to restrict amplification of otherwise lawful content.

The Right to Be Forgotten effectively created a new notice-and-takedown legal regime, with the same over-enforcement risks and procedural questions we saw under illegal speech models. But it also established a whole new legal category of restricted speech. Defining the substantive scope of this new prohibition required answering hard questions. How many years must pass before a convicted drunk driver or domestic abuser is entitled to hide his past from search results? Does it matter if he is currently employed as a

driver or a childcare provider? (As counsel for Google at the time, I helped make decisions like these.) Older cases or laws regarding publishers might seem to answer some of these questions, but *Google Spain* made clear that this precedent didn't apply. Delistings from web search should cover all the information that was already *illegal* to publish, and also an additional margin of information that was *legal* to publish.

In the years since the ruling, a small body of public case law has developed—as well as a much larger body of internal precedent, which Google is barred from disclosing in detail.[54] The case law exists because Google asserted a public interest basis to reject many delisting requests, and litigated those cases before regulators and courts. That's a costly investment. It is not one the public can count on platforms to make for most laws—particularly not laws governing social media platforms, which typically want to take down lawful but policy-violating content anyway. Search engines, by contrast, have a financial interest in offering comprehensive results. A harmful speech model without a mechanism for public review of disputes raised by affected speakers can leave the definition of new and untested rules entirely in the hands of platforms.

More recently, some European countries have moved toward potential restrictions of "harmful" but lawful speech by extending or building on existing media or communications regulation. The EU's 2018 updates to the Audiovisual Media Services Directive give existing national media regulators authority over video platforms like YouTube.[55] The U.K.'s pending Online Harms legislative package will empower its media regulator to oversee platform content moderation generally, as will pending Irish regulation.[56] The U.K.'s plan includes requiring platforms to take down "harmful" but previously lawful content. Like the Right to Be Forgotten, it will establish a new and thus-far-undefined category of restricted online speech. Unlike the Right to Be Forgotten, it will not involve the regulator in resolving any individual takedown decisions.

European legal cultures' greater willingness to trust regulators and greater tolerance for restrictions on expression makes the adoption of such new restrictions more feasible. In continental civil law systems, too, it is more normal to rely on statutes or regulations rather than case law. But under any legal system, setting forth new speech rules without giving speakers a clear

path to challenge them is very problematic. New rules will always prove hard to apply in difficult or unexpected real-world cases. The resolution affects not only speakers' and listeners' rights, but the public's interest in clearly defined rules.

Regulators' reluctance to field individual disputes is understandable given the sheer volume. Google has so far assessed some 4 million web pages under *Google Spain*.[57] Facebook has been known to take action against 2 *billion* items of content in a six-month period. No regulator in the world has the capacity to provide nuanced analysis of all those questions.[58] But in the absence of regulatory capacity—or even more expensive judicial review—it is hard to identify a sufficiently rights-respecting harmful speech model, even if one accepts the need to restrict additional speech online or in amplification features.

B. U.S. law and communications law precedent

The idea of categorically banning certain kinds of currently legal speech from internet platforms, or from their amplification features, will sound alien to most U.S. First Amendment lawyers.

U.S. law does not provide much precedent for holding whole categories of speech legal in one place, but illegal in another. And distributors of speech usually face less liability than speakers do—not more.

But the harmful content model bears examination, not least because so many lawmakers have loudly called on platforms to restrict the reach of lawful speech in the U.S. In any case, our jurisprudence isn't blind to context and consequences of speech. It matters where and how words are used. Under the seminal *Brandenburg* "incitement to violence" standard, for example, words that are permitted in one's home may become illegal when spoken to an angry mob. Lyrissa Lidsky and others have closely examined how this rule might—or might not—adapt in light of the magnified risk posed by online incitement, given its changed reach, permanence, anonymity, and social context.[59] Notably, at least one pre-*Brandenburg* case, *Dennis v. U.S.*, appears to limit amplification of specific messages. Defendants were convicted for teaching other people about Communist literature—although the literature itself was, in printed form, legal.[60] These cases, though, address very specific instances of speech in claims against the speakers themselves.

They do not do much to support categorical restrictions for platforms or their amplification features.

The closest precedent for the harmful speech model in the U.S. comes—much as it does in Europe—from communications law. Lawmakers responding to earlier communications technologies, like lawmakers now, worried that legal but previously hard-to-find content had become more discoverable, and thus more dangerous. Some of the laws they passed to restrict content on broadcast or cable survived First Amendment challenges—making them highly relevant to today's discussion about regulating amplified online content. The *Playboy* decision mentioned above, for example, did not reject the premise that Congress might have a special interest in regulating pornography on cable TV. Rather, it concluded that the regulation in that case did too much collateral damage to lawful speech, and thus was not narrowly tailored enough to meet First Amendment requirements.

In *FCC v. Pacifica*, by contrast, the Court upheld media-specific restrictions.[61] The FCC, it held, could penalize a radio station for broadcasting a profanity-laden comedy routine that would have been legal in other venues, like comedy clubs. Uniquely restrictive speech rules for broadcast were justified, the Court reasoned, given radio's ability to intrude unexpectedly into the home of an unwilling listener, and given the potential presence of children in the audience. The Court has also upheld laws requiring broadcast and cable companies to carry *more* content than their operators wished to, in service of goals like media pluralism and competition.[62] In the *Turner* cases, it recognized Congress's authority to override cable companies' editorial discretion—including what we might now call "amplification" decisions to allocate the best channel numbers to cable companies' preferred programming.[63]

Applying broadcast or cable precedent to support online amplification restrictions would be very hard, though. The U.S. Supreme Court emphatically rejected analogies to those cases in *Reno v. ACLU,* the leading case examining and striking down internet speech restrictions. The "special justifications for regulation of the broadcast media," including the scarcity of available frequencies, the Court said, were "not present in cyberspace."[64] Some of *Reno*'s reasoning, including that the "risk of encountering indecent material by accident is remote because a series of affirmative steps is required to access specific material," arguably might not apply to features

like recommendations on today's platforms.[65] And certainly the concentration of online speech on platforms and ranking systems controlled by a few large companies were not factors in *Reno*. Still, a congressional attempt to restrict online amplification of lawful speech in the face of *Reno* would be a very heavy lift.

A communications law-based rule restricting amplification of harmful content in the U.S. would also raise the same administrative issues it has in Europe. Someone—presumably the FCC—would need to articulate new speech rules, analogous to the ones currently under development in the U.K. Those rules could not simply replicate older standards, crafted for the tiny cadre of privileged 20th century speakers who amplified their speech via TV or radio. Rules developed for Dan Rather or "Diff'rent Strokes" will not be the right fit for today's multitudinous cacophony of online speech. Nor should they govern ordinary people's daily communications, which populate many ranked social media feeds. It is hard to imagine the FCC—which has enforced its indecency rules only a handful of times in the past decade—working to set those rules, and adjudicating the resulting tidal wave of disputes.[66] If it doesn't, though, then the job of refining untested new speech rules will either sit with platforms, or inundate the courts.

Given *Reno*'s analysis, the odds seem low that a U.S. law following the harmful speech model would survive constitutional review. Chances that Congress will build up the necessary regulatory infrastructure seem lower still. That brings us to the final set of models: content-neutral rules that avoid the need to legally define restricted speech.

IV. CONTENT-NEUTRAL MODELS: INCREASING PLATFORM LIABILITY FOR AMPLIFYING ANY SPEECH

A **FINAL SET OF MODELS** would restrict amplification without reference to specific prohibited content. Lawmakers, particularly in the U.S., might consider such a content-neutral approach their best option to avoid the more stringent scrutiny that courts apply to content-based restrictions on speech. Indeed, as discussed below, some of the biggest legal

questions about proposals in this category involve whether a particular rule actually achieves content neutrality.

In this section, I will first examine the idea of eliminating amplification entirely. That model is hard to assess, because it is hard to define. I will then discuss two content-neutral rules that strike me as relatively promising: "circuit breakers" to slow the spread of highly viral content, and laws to give internet users more control over algorithmically ranked content using competition, privacy, or possibly consumer protection rights. For each, I will sketch some general observations about First Amendment issues—though my conclusions on these more novel questions are relatively tentative, compared to the discussions of case law so far.

A. Getting rid of amplification

Lawmakers could, in theory, make platforms "turn off" their algorithms, prohibiting amplification altogether.[67] Or they could do so for just some products or features, to avoid disabling ranking-dependent services like search engines. More realistically, they might accomplish something similar through aggressive liability or regulation under the illegal or harmful content model rules discussed above. A legal message that said "if you can't avoid amplifying illegal content, then don't amplify anything at all" would surely lead some companies—especially those less economically equipped to assume legal risk—to choose the latter.

To draw conclusions about that trade-off, we need to be clearer about what the resulting platform services would actually do. What counts as "amplification," and what do we believe an authentic, un-amplified service looks like? Depending on our answers to those questions and our own policy priorities, we may have different ideas about which platform design choices are legitimate and authentic, when or why algorithmic intervention is warranted, and whether we would be better off without those algorithms.

1. CHRONOLOGICAL ORDER AND DESIGN DEFAULTS

One common proposal would require or incentivize services like social media to show user posts in reverse-chronological order—putting the newest posts at the top of a newsfeed.[68] That would effectively eliminate the value of features like search results or YouTube recommendations tailored to a user's

interests. But for things like the Twitter or Facebook newsfeed, it could also eliminate many possibilities for mischief in platforms' ranking choices—in exchange for forfeiting the positive sorting and discovery functions of ranking.[69] It would also have a number of more subtle downsides.

One problem has to do with repetition. A chronological newsfeed can be spammed by people or bots posting the same or similar things every few seconds. That's somewhat less of a problem if spam violates a clear rule and can just be removed—though platforms may still want to do things like reducing a post's visibility while it is queued for review by moderators. Spam is also less of a problem for users who see posts only from known and trusted accounts. Even that restrictive set-up can create problems, though. If 50 people I follow on Twitter all retweet the same post, should I see that post 50 times, or one time? Showing a popular post just once would obscure the other 49 people's posts. (Or replace them with some kind of duplicate count, which would also depart from strict chronological order.) Showing the popular post 50 times would avoid putting Twitter's thumb on the scale, but it would be a bad user experience.

Another problem has to do with human behavior. There is reason to believe that a purely chronological system will show more "borderline" content—material that *almost*, but not quite, violates whatever speech prohibitions a platform enforces. According to Facebook, no matter what line the company draws in prohibiting content like misinformation or racist language, users are incentivized to create and post a high volume of material that comes close to crossing the line without quite doing so. There is no reason to assume this problem is unique to Facebook. Rather, as Mark Zuckerberg put it in a blog post, it likely reflects a "basic incentive problem" for people to create provocative content because it gets more reactions from other platform users. (And while this user behavior in turn may prompt Facebook to boost content in its own ranking, that cycle doesn't start with Facebook. It starts with users' own behavior and apparent preferences, and would exist in some form without Facebook's algorithmic intervention.) Facebook currently addresses this seemingly unavoidable glut of borderline content through demotion—ensuring that material in this category "gets less distribution and engagement."[70] In a purely chronological system, that would not be an option. Users would see more barely legal or barely terms-of-service-compliant content.

Facebook's experience with "borderline" content illustrates another issue: Our ideas of authentic platform operation may well be artifacts of choices made by platforms themselves. The entire category of "borderline" content is an example, because it is defined by how close it comes to violating Facebook's Community Guidelines.[71] The glut of borderline posts in a chronological feed would be driven by user behavior, but the *kind* of content being shared would be a byproduct of Facebook's own rules. If the platform prohibited full nudity, we would see more partial nudity, if it prohibited real-world violence we would see more simulated violence, and so forth.

At a macro level, if we define and mandate authentic platform operation based on current incumbents' models, we may miss out on future competitors and technologies to better address our problems. At a micro level, defining expectations based on platforms' own choices is just hard to get away from. One illustration comes from the Knight First Amendment Institute's litigation about President Trump's Twitter account. Plaintiffs in that case successfully argued that the president deliberately embraced Twitter's open-access default—and thus could not block users from his account without violating the First Amendment. Trump's lawyers, meanwhile, tried to position user control over Twitter's blocking function as the default and authentic product behavior. According to them, plaintiffs themselves were trying to override the platform's normal behavior, and "claiming a right to 'amplify' their speech by being able to reply directly to the President[.]"[72]

2. AUTHENTIC AND INAUTHENTIC PATTERNS OF USER BEHAVIOR

The difficulty of defining platforms' authentic and unamplified base state is compounded by internet user behavior. On some services, like text messaging apps, users themselves may promote harmful content by sharing it widely. On other services, platforms may draw on aggregate patterns of user behavior to shape recommendations and rankings. Some, like Reddit, adopt simple up/down voting systems. Others rank algorithmically based on logged data about user behavior. (I'll talk more about that later in this essay.) These systems all share an underlying problem: People on the internet are terrible. Users often share misinformation deliberately. Ranking based on votes can produce a "garbage in, garbage out" phenomenon, in which popular but harmful content like "Tide Pod Challenge" videos can rise to the top. So can

racist, misleading, or privacy-invasive content. Platforms can accurately reflect that too-real pattern or they can intervene—sacrificing authenticity for other values, like child safety or racial equality.

Ranking based on user votes or other activity also opens the door to "coordinated inauthentic behavior," including campaigns to boost misleading or harmful content. Versions of this problem arise on nearly every internet platform, using ever-evolving techniques. Reddit, for example, reports that what it calls "content manipulation," which includes both commercial spam and politically motivated "brigading," occupies the majority of its content moderation efforts.[73] Content that gets promoted by these means is sometimes so undesirable that platforms remove it altogether. In other cases, the content itself is innocuous, but showing up in the wrong place. That happened, for example, when a "googlebombing" campaign caused the official White House home page to appear at the top of search results for the term "miserable failure."[74] The White House page wasn't bad per se. It was just badly amplified.

Distinguishing the useful signals of crowd behavior from the misleading signals of coordinated inauthentic behavior is not necessarily easy, either in theory or in practice.[75] As evelyn douek puts it, "Coordination and authenticity are not binary states but matters of degree, and this ambiguity will be exploited by actors of all stripes."[76] At the edges, defining the authenticity of coordinated behavior like boycotts, petition drives, or hashtag-based advocacy like #metoo may unavoidably require value-laden judgment calls.

If we define amplification to include the effects of coordinated "inauthentic" user behavior, then getting back to a more "authentic" baseline requires platform intervention, including through adjustment of algorithms. If amplification means platform manipulation of ranking, then our "authentic," un-amplified platforms will be plagued by coordinated "inauthentic" user behavior. Defining amplification is hard. Defining it in ways that serve policymakers' goals without creating other unintended consequences is likely to be harder. I have largely assumed these definitional problems away in assessing the models so far. (Otherwise, engaging in the policy discussion about amplification would be impossible.) But the problems exist, and the idea of eliminating amplification entirely, or penalizing platforms that do it, puts them in stark relief.

3. FIRST AMENDMENT CONCERNS

However we define amplification, a rule that prohibited it would face potential First Amendment problems involving the speech rights of both users and platforms. A deep dive into these questions lies beyond the scope of this essay, because such a drastic legislative option strikes me as largely hypothetical. Here is a sketch of the issues, though.

A. USER RIGHTS

Some users would almost certainly perceive harm to their speech rights if they lost access to a large audience because the government shut down platform amplification. Lawsuits brought on that basis would look something like right-wing commentator Dennis Prager's case against YouTube. Prager argued that YouTube had violated his First Amendment rights by demonetizing and otherwise disfavoring his videos. The Ninth Circuit firmly rejected that claim, because YouTube is not a state actor and thus not bound by the First Amendment.[77] But if Prager suffered the same harms from actual state action, his lawsuit would not have that problem.

I'm not sure how a user's First Amendment challenge to blanket laws restricting amplification might ultimately end, though. On one hand, users clearly can bring First Amendment claims when a law causes platforms to police speech, including amplified speech, too aggressively. So it would seem perverse to give them no recourse when a law causes platforms to terminate amplification entirely. And users would likely argue that such a law's purported content-neutrality was a pretext, given many lawmakers' oft-stated goals of restricting specific content. Such a law would also arguably burden users' rights to free assembly—particularly if, in a time of pandemic and social distancing, the state effectively reduced users' ability to share their own messages widely among friends, like-minded activists, or constituents. On the other hand, the state shuts down potential forums for speech, like unruly or unhygienic taverns, all the time. That kind of state action, driven by content-neutral goals like public health, is clearly permitted. So are "time, place, and manner" restrictions that deny all speakers the right to use megaphones, or to do so at particular times. If a court accepted that a ban on platform amplification was similarly content-neutral, perhaps users' objections on First Amendment grounds would fail.

B. PLATFORM RIGHTS

The other First Amendment claim would be that of platforms themselves. As discussed above, platforms' own algorithmic ranking and recommendation have been held to constitute protected speech. A law explicitly prohibiting such speech—or requiring platforms to replace their own preferred algorithm with the state's preferred algorithm, as a chronological ranking mandate would do—is likely to face real constitutional problems. It would not, I think, be content-neutral as to platforms. For platforms (unlike users), it would rule out specific messages. That includes very generic messages such as "I predict you'll like this," but also more value-laden messages expressed through up- or down-ranking of particular medical advice or news sources.[78] Such a restriction would, among other things, reshape the balance of corporate power between platforms and traditional media companies.[79] CNN and Fox News would remain free to promote the messages of their choice, by the same means they do now. Platforms would not.

The complete legal elimination of platform amplification strikes me as unlikely. The more plausible law, making amplification so legally risky that platforms eliminated it themselves, would face the separate issues of the illegal speech models discussed in Section II. Some potentially more realistic content-neutral approaches to amplification, however, are spelled out in the remaining two subsections.

B. Circuit-Breakers

One possible non-content-based law would be a "circuit-breaker" rule, permitting amplification only up to some quantified limit. That limit might be defined by metrics like the number of times an item is displayed to users, or an hourly rate of increase in viewership. Like older laws restricting use of physical amplifiers or loudspeakers, such a rule would in principle turn on a message's *volume* rather than its *content*.[80] A loose legislative model might be the 1992 Audio Home Recording Act, which limited the number of high-fidelity copies made by digital audiotape devices.[81]

A circuit-breaker rule could potentially trigger more complex consequences than a flat ban on further amplification. For example, platforms that amplify information past a certain high level of virality might be considered "on notice" of it, and lose legal immunities for readily identifiable

illegal content. That might prove beneficial—or it might push platforms to misallocate resources, spending time assessing benign social media crazes at the expense of more serious threats. Such a law would also risk replicating the over-enforcement problems of illegal speech models. In any case, circuit breaker rules are not a cure for all of the problems with amplification—only the ones that create spikes in usage patterns. For the narrow but important situations in which clearly illegal content is spreading like wildfire, though, they might provide a firebreak.

1. USER RIGHTS

It's not entirely clear that a circuit-breaker rule would have neutral impact on user speech, though. Breaking news, for example, is particularly likely to cause sudden spikes in user engagement and interest, with resulting amplification by platforms. So novel or newsworthy posts, including extremely important material like the videos documenting the deaths of Philando Castile or George Floyd, would be disproportionately affected by a cap on amplification. For similar reasons, circuit-breaker rules would effectively favor certain speakers over others. A rule like "no one can have more than X followers on social media" might disfavor popular accounts, effectively benefiting less popular ones. A rule like "no account may add new followers at a rate of more than 10 percent each day" would favor more established voices at the expense of marginalized ones. Restraints on abruptly popular voices or messages would effectively keep the microphone in the hands of whoever already has it—preserving the status quo at the expense of little-known artists, activists, and others hoping to "go viral."

Some of these concerns about content-neutrality of amplification rules go back at least to mid-20th century cases about mechanical amplification. In a 1949 case upholding an ordinance against sound trucks, for example, the dissent argued that the ruling would "give an overpowering influence to views of owners of legally favored instruments of communication," and "preference in the dissemination of ideas" to those who "obtain the support of newspapers" or other established media companies.[82] The Court's disregard for such concerns back then might bode ill for users raising similar objections to a circuit-breaker rule now.

2. PLATFORM RIGHTS

A circuit-breaker rule would also, of course, affect platforms' own First Amendment rights. Like an outright ban, it would stop platforms from communicating recommendations—it just wouldn't do so nearly as frequently. And like a ban, it would have some consequences for the relative roles of platforms and traditional media in the information ecosystem. A narrower circuit-breaker rule that dampened only rapidly spreading content would particularly restrict platforms' ability to "speak" about issues like breaking news.

C. Privacy and competition laws

Finally, options based not on speech regulation but on privacy, competition, or possibly consumer protection may offer some paths forward, bypassing many of the constitutional difficulties described above.[83] These models are, at heart, about increasing user agency. They would not prevent individuals from actively choosing lawful but harmful or polarizing content online, any more than current law prevents that same choice in media consumption. But by increasing users' options, and ideally increasing the diversity of platform amplification offerings, these approaches could alleviate other important problems relating to online content.

Algorithmic ranking systems typically draw on aggregate patterns within data sets reflecting human behavior, in order to predict what content users will want to see in new situations. Some of this data, like the website links that power Google's PageRank algorithm, is public. Other data, like logs of individual users' clicking and browsing behavior, is private. Logged data can be used in anonymized or aggregate forms, and used to rank results and recommendations that all users see. An algorithm might prioritize a certain result for every search for "dog," for example, because past users who searched for "dog," in aggregate, clicked most on that one. Logged data can also be used in non-anonymous form, and shape personalized results, recommendations, or newsfeeds that vary for each user. YouTube might recommend new "Saturday Night Live" skits because I watched them in the past, for example. Or it might know that I have watched Sister Rosetta Tharpe performances, and also know that users who watch blues videos, in aggregate, tend to watch jazz videos. As a result, it might recommend footage of Cab Calloway to me.

A common critique of systems that amplify or target content based on

user behavior is that, in conjunction with ads-based business models, they cause platforms to amplify anything that keeps users engaged, including harmful, misleading, or extreme content. As an EU Parliament report put it, a platform that optimizes for ad revenue has reason to prioritize "content based on addressing emotions, often giving rise to sensation in news feed and recommendation systems[.]"[84] As a result, users who start out interested in center-right politics might, for example, see recommendations for extremist conspiracy theories.[85]

To my mind, this theory's focus on behavioral data tracking is sound, but I suspect its emphasis on advertising is somewhat overstated.[86] After all, people choose sensationalist or provocative content when they have to pay for it, too. That's why grocery stores position gossip magazines in the checkout aisle for impulse purchase. And people using the internet often amplify engaging but unreliable content on platforms with no ads, and no algorithmic ranking. That's why 1990s email users received so many dubious political chain emails from gullible relatives. It's why electoral disinformation and other harmful untruths go dangerously viral on platforms like WhatsApp, which has no ads and no platform-initiated ranking. As an EU official pointed out about Parler, which had not ramped up its advertising business when it became a common vector for misinformation, "banning advertising in that case … would not have changed the virality."[87]

To be clear, there are plenty of reasons to regulate online advertising. I think the deeper questions about behavioral targeting and amplification, though, are not about ads. They're about ranking based on our online behavior, when—as mentioned above—humans on the internet are terrible. Our collective behavior seems to reveal a real demand for junk—often including disturbing, bias-affirming, or outrage-generating material. Platforms keep showing us sensational content because we—individually or in aggregate—keep clicking on it.

If our behavior—our "revealed preferences," in economic parlance—says we want trashy but legal content, should laws prevent platforms from giving it to us? Those who answer "yes" to this question will presumably be interested in the harmful speech models for restricting amplification of currently lawful speech, as discussed in Section III. Those who answer "no," but who still wish platforms wouldn't promote so much garbage, are better

off focusing on approaches that are not grounded in government restrictions on content.

Online behavioral data can show human beings at our worst. The rapid-fire clicking and browsing that platforms track are shaped by the same minimally evolved parts of our brains that prompt us to stare at traffic accidents, grab gossip magazines in the checkout line, or use racist stereotypes as cognitive shortcuts.[88] The data that platforms collect may also reflect what we do when we think no one is watching. Notoriously, that is when we are least likely to do the right thing from a societal perspective, or honor social norms.[89]

Many of us would make different choices about online content consumption if we had the opportunity to consult our better selves—to bypass our lizard brains and use our more developed cerebral cortexes, our superegos, our faiths, or whatever else we draw on for conscious and value-driven decisions. The premise that our clicking and browsing behavior does not necessarily reflect what we *really* want is useful, I think. It opens up space to argue for healthier intellectual fare as a matter of user autonomy, rather than as a top-down restriction on speech and information. It also lets us look to other sources of law, besides speech regulation, to address problems with amplification. Major sources of applicable law are privacy, competition, and potentially consumer protection.

1. PRIVACY

Statutory mandates grounded in protection of user privacy could let us choose less extreme material, by giving us more granular control over how our personal data is used to target content. Settings could allow us to exclude collection of certain data, or to exclude how that data is used—including for targeting via algorithmic ranking systems. We could dial down political snark, and dial up cat videos or poetry or history podcasts. Perhaps we could even choose which behavior platforms track and use—preventing targeting based on things we clicked on too quickly, or too late at night, or too soon after texting with our ex-boyfriends. Modest steps in this direction already exist in features prompting users to slow down, like the warnings Twitter shows to users who retweet articles without clicking to read them first,[90] or Facebook's user settings for newsfeed content ranking.[91] Somewhat larger

steps will come from the EU's pending Digital Services Act, which will require larger platforms to let users turn off personalized recommendations, and show them any options to adjust the parameters used in ranking.[92]

A. USER RIGHTS

This approach avoids many First Amendment problems, because it does not involve government preferencing of content. Internet users who lost reach and audience under such a law would have little basis to object, since that loss would stem from other users' choice not to listen to them. Indeed, the Supreme Court's examples of narrower tailoring in prior cases about communications technologies often involved exactly this: letting individuals decide what content they want to see, rather than putting that decision in the hands of an intermediary or other centralized authority.[93]

B. PLATFORM RIGHTS

Platforms would have some First Amendment arguments against privacy-based controls on amplification, though. The government would be reducing the companies' current leeway to determine algorithmic ranking—not to the same degree as it would by banning amplification entirely, but in a way that could prompt similar legal objections based on platforms' First Amendment rights to set editorial and ranking preferences. I'd like to think that major platforms, including my former employer, would not participate in such suits. But showing users the staid or highbrow fare they claim to want, instead of the emotionally engaging material they have actually opted for in the past, would almost certainly be bad for platforms' bottom line—regardless of their revenue model. For major incumbent platforms, a serious change rooted in expanded user privacy rights (as opposed to the modest changes some have made voluntarily) could be far more threatening than the illegal or harmful speech models discussed above.[94] Companies that can afford to adapt to content-based liability—by expanding already-enormous content moderation teams, "voluntarily" excluding large categories of content from the platform or its amplification features, and sometimes litigating or settling cases—may prefer that to systemic business model change.

2. COMPETITION

Changes grounded in competition could also ease many of the problems with amplification today. In a world with dozens of competing social media platforms or search engines, the entire issue of amplification would be different. No single platform would, in principle, shape the information diet of such enormous audiences. So the "reach" of any given platform's recommendations or ranking would be much reduced.[95] Having a diversity of competing platforms would also, like the privacy approach described above, let individuals choose less politicized options—like switching from The Daily Beast to Newsweek. Any legal change along those lines would be a massive undertaking from a competition law perspective, of course, and face fierce resistance from large platforms. But neither their First Amendment rights, nor those of users, would provide a major part of their legal arsenal.[96]

Many of the biggest challenges to these approaches—other than those grounded in competition law itself—would be practical. Platforms like social networks or messaging apps are notoriously subject to network effects. The more users they have, the better their service is for each user. As a result, it is generally assumed that if a platform like Facebook were broken up, economic forces would drive it back together. Today's dominant platforms also sit on a dragon's hoard of user data and content, which would be all but impossible for newcomers to replicate.

But the law is not without tools for problems of this sort. One comes from "unbundling" requirements. In telecommunications law, these function to insert competition into markets subject to network effects by requiring incumbents to license hard-to-duplicate resources to newcomers. Various thinkers have floated proposals of this sort for platforms, under monikers including "protocols not platforms," "Magic APIs," "algorithmic choice," or "middleware."[97] This approach would allow competitors to build on data and content held by incumbent platforms, using it to offer users different ranking algorithms to choose from, as well as different content moderation rules or even user interface design. Users could opt for their church's ranking preferences, or Vox's, or Fox News's—or even just Facebook's, Google's, or Twitter's. In some versions of this model, removal of illegal content could still be done by the platform at the "hub" of this system, and propagate out to competing "spoke" services.

An undertaking like this would be very, very complicated. It would require lawmakers and technologists to unsnarl many knots, as I have discussed elsewhere.[98] But unlike many of the First Amendment snarls described above, these ones might actually be possible to untangle. Technologist Stephen Wolfram testified to Congress about how to address the technical challenges in 2019.[99] Twitter has backed a project to make distributed ranking and moderation a reality—with stated goals including "keep[ing] controversy and outrage from hijacking virality mechanisms."[100] Experts in Europe and elsewhere are wrangling with ways to protect user privacy in interlocking or interoperable services.[101]

Of course, competition or privacy solutions that protect user autonomy are not helpful if we are all worried that other people will choose the wrong content and ranking priorities. Improved user choice will not help with echo chambers or filter bubbles. But it may address enough other problems to make those ones seem less pressing, or more possible to tackle by other means. In any case, overriding users' choices and forcing them away from their preferred diet of lawful-but-awful content is not a project to be undertaken lightly, or with much optimism about success in the face of First Amendment challenges.

3. CONSUMER PROTECTION

Consumer protection law—a term I'll use broadly to encompass things like food and drug regulation—is a major source of legal precedent for telling businesses *not* to give people what they want or are willing to buy. That makes it an interesting model for addressing issues of harmful but legal online content.[102] I am less optimistic about it, though.

Consumer protection laws restricting amplification of particular content would face the same First Amendment challenges as any other illegal or harmful speech model, as discussed in Sections II and III. We may speak figuratively about avoiding "intellectual junk food" or making users "eat their veggies" by consuming a more diverse or healthy news feed.[103] But speech is not food. Congress has far less power to regulate our intellectual diets than our corporeal ones. As the Supreme Court put it in *Smith*, "[t]here is no specific constitutional inhibition against making the distributors of food the strictest censors of their merchandise, but the constitutional guarantees of

the freedom of speech and of the press stand in the way of imposing a similar requirement on the bookseller."[104] Analogies to drug regulation, or even gambling regulation, seem unavailing for much the same reason.[105]

That leaves the more anodyne application of consumer protection law: requiring platforms to accurately "label their goods," by explaining their algorithms. This approach appears in several recently proposed laws. Part of the idea is that with better information, users can make better choices. Of course, without more competitors to choose from, those choices may not be so meaningful. As a First Amendment matter, labeling requirements would not use state power to force changes to ranking algorithms, so they would in that respect be content-neutral. They would, however, compel platform speech in the form of labels or explanations. At least one appellate court has struck down a similar requirement on First Amendment grounds.[106]

To my mind, better transparency about the parameters of major platforms' algorithms would benefit consumers, lawmakers, and society at large. On its own, though, such transparency would do little to counter amplification of harmful or illegal content. That, combined with vulnerability to First Amendment challenges, makes this approach seem ultimately less useful and interesting than those grounded in privacy or competition.

V. CONCLUSION

Current debates about regulating amplification have a lot in common with long-running discussions about regulating content moderation in the first place. In both cases, the rules that platforms apply now have been the subject of extensive and justified criticism. In both cases, requiring platforms to be "neutral" or to refrain from exercising judgment about content would make many users' experiences worse, and open the door to ugly and harmful material.

Laws assigning platforms liability for hosting content and laws assigning liability for amplifying it also share practical issues. Someone has to decide which content is excluded, whether it is removed from the whole platform or just from features like recommendations. In an illegal speech model, that choice would be made by a machine or platform employee trying

to apply the law. A harmful speech model would be the same, except that someone would also have to define the scope of new speech restrictions. In either case, platforms' safest course would be to suppress *more* speech than lawmakers intended. Such errors in demoting content, like errors in deleting it, can be checked by better processes within the platform or before courts or administrative agencies. But the better those processes are, the greater their cost, for the platform and for any agencies or courts tasked with resolving disputes.

And of course, for both ordinary content liability laws and proposed laws about amplification, the First Amendment plays a structuring role. Congress cannot go too far in requiring or incentivizing platforms to take down legal speech. The same constitutional limits apply if Congress wants platforms to demote or cease amplifying that speech. This means that content-based rules about amplification, whatever we think of their wisdom, are very difficult to craft. The high risk of running into constitutional dead-ends provides one of many reasons to explore options grounded, instead, in privacy or competition.

There are a few relevant issues I have not examined in this already-long essay. One is the policy logic behind a rule that gives platforms freedom to remove content but not to demote or arrange it. Do we want to leave platforms with only the clumsiest tools for responding to problematic content—or for serving policy goals like featuring diverse voices, promoting authoritative news sources, or simply providing information that users want and can use? Another is the idea—more visible in current European discussions—that lawmakers might simply dictate ranking criteria based on policy goals.

Others are more specific to U.S. legal doctrine. For example, does First Amendment analysis change when platforms do not amplify specific user posts, but instead nudge users to connect with particular groups or accounts? That avoids penalizing particular speech, but at the cost of preemptively penalizing particular speakers—a move that, in other contexts, would be an impermissible prior restraint. Does Congress have more leeway to act against lawful speech if it does not *require* platforms to impose restrictions, but instead makes them the condition for statutory immunities like those under CDA 230? I think it does not.[107] But applicable law under the unconstitutional conditions doctrine, which the Supreme Court has called "notoriously tricky," deserves more careful attention.[108]

Questions about the First Amendment and amplification are tricky, too. I hope to see more careful work in this area, which I expect will only become more important as platform content moderation practices evolve. Harms attributable to amplification are real. The constraints created by First Amendment law are also real. Any paths forward will require precision and realism about both.

NOTES

1 *See* Joshua A. Tucker et al., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (Hewlett Found. Mar. 2018), https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf [https://perma.cc/R93T-CJTM].

2 Zeynep Tufekci, *YouTube, the Great Radicalizer*, N.Y. Times (Mar. 10, 2018), https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html [https://perma.cc/ZV4S-ZL6V].

3 Casey Newton & Zoe Schiffer, *What Facebook should do about its Kenosha problem*, Verge (Sept. 1, 2020, 6:00 AM), https://www.theverge.com/interface/2020/9/1/21408650/facebook-kenosha-guard-policy-moderation-public-report. This narrative is not without its critics, including careful researchers who question the role of platform algorithms in the increased popularity of far-right content on YouTube. Paris Martineau, *Maybe It's Not YouTube's Algorithm That Radicalizes People*, Wired (Oct. 23, 2019, 7:00 AM), https://www.wired.com/story/not-youtubes-algorithm-radicalizes-people [https://perma.cc/7MX8-7639]; Kevin Munger & Joseph Philips, *Right-Wing YouTube: A Supply and Demand Perspective*, Int'l J. Press/Pol. (Oct. 21, 2020), https://journals.sagepub.com/doi/abs/10.1177/1940161220964767 [https://perma.cc/H77Z-6DL9].

4 I have argued elsewhere that discrimination claims like this are likely not properly subject to intermediary liability immunities under CDA 230. Daphne Keller, *Toward a Clearer Conversation About Platform Liability*, Knight First Amend. Inst. Colum. Univ. (Apr. 6, 2018), https://knightcolumbia.org/content/toward-clearer-conversation-about-platform-liability [https://perma.cc/KHX9-EHB4].

5 Safiya Noble, *Google Has a Striking History of Bias Against Black Girls*, Time (Mar. 26, 2018, 4:30 PM), https://time.com/5209144/google-search-engine-algorithm-bias-racism/ [https://perma.cc/U4DM-3DFG].

6 Natasha Lomas, *Google fined $2.7BN for EU antitrust violations over shopping searches*, TechCrunch (June 27, 2017, 5:54 AM), https://techcrunch.com/2017/06/27/google-fined-e2-42bn-for-eu-antitrust-violations-over-shopping-searches/ [https://perma.cc/F7XC-SJ8A].

7 For more thorough discussions of the underlying technologies and policy issues of amplification, see Emma Llansó et al., *Artificial Intelligence, Content Moderation, and Freedom of Expression* (Feb. 26, 2020) (unpublished working paper), https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf [https://perma.cc/VLK8-B5W6] (discussing algorithmic curation and human rights); Spandana Singh, *Why Am I Seeing This?* (Open Tech. Inst., Mar. 25, 2020), https://www.newamerica.org/oti/reports/why-am-i-seeing-this/ [https://perma.cc/ZS7C-7EX7] (discussing recommendations); Spandana Singh, *Rising Through the Ranks* (Open Tech. Inst., Oct. 21, 2019), https://www.newamerica.org/oti/reports/rising-through-ranks/ [https://perma.cc/N7UY-845S] (discussing search results and newsfeeds); Paddy Leerssen, *The Soap Box as Black Box: Regulating Transparency in Social Media Recommender Systems*, 11 Eur. J. L. Tech. (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544009 [https://perma.cc/WR3Z-87PE] (discussing models for transparency and accountability of social media recommendation algorithms). For a speech-law-based discussion, see Erin L. Miller, *Amplified Speech*, 43 Cardozo L. Rev. (forthcoming 2021). For a review of tools currently used to demote or reduce viewership of particular content, see Eric Goldman, *Content Moderation Remedies*, 28 Mich. Tech. L. Rev. (forthcoming 2021).

8 These arguably exist on a continuum, with search results increasingly trying to anticipate users' needs, and newsfeed results responding to a user's implicit request to be shown something interesting.

9 *Google search changes tackle fake news and hate speech*, BBC (Apr. 25, 2017), https://www.bbc.com/news/technology-39707642 [https://perma.cc/69AN-ZKDN].

10 Llansó et al., *supra* note 7, at 3.

11 Section 230 of the Communications Act of 1934, File No. RM-___(Nat'l Telecomm. and Info. Admin. July 27, 2020) (petition for rulemaking), https://www.ntia.gov/files/ntia/publications/ntia_petition_for_rulemaking_7.27.20.pdf [https://perma.

cc/Y6DS-F9Z8].

12 *Draft Report with recommendations to the Commission on a Digital Services Act 2020/2019(INL) by Committee of Legal Affairs*, at 23 (Apr. 22, 2020), https://www.europarl.europa.eu/doceo/document/JURI-PR-650529_EN.pdf [https://perma.cc/GL9C-4XPW].

13 Mackenzie Nelson & Julian Jaursch, *Germany's new media treaty demands that platforms explain algorithms and stop discriminating. Can it deliver?*, Algorithm Watch (Mar. 9, 2020), https://algorithmwatch.org/en/new-media-treaty-germany/ [https://perma.cc/SHJ6-EB68].

14 Code of Practice on Disinformation, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation [https://perma.cc/8CD5-VZWW] (last visited May 26, 2021). See analysis of this conflict in *Germany proposes Europe's first diversity rules for social media platforms*, London Sch. Econ. Media@LSE Blog (May 29, 2019), https://blogs.lse.ac.uk/medialse/2019/05/29/germany-proposes-europes-first-diversity-rules-for-social-media-platforms [https://perma.cc/V5CF-C874].

15 Mark Zuckerberg, *Preparing for Elections*, Facebook (Mar. 13, 2021), https://www.facebook.com/notes/mark-zuckerberg/preparing-for-elections/10156300047606634/ [https://perma.cc/KHG8-FRKW]; Tessa Lyons, *Hard Questions: How Is Facebook's Fact-Checking Program Working?*, Facebook (June 14, 2018), https://about.fb.com/news/2018/06/hard-questions-fact-checking/ [https://perma.cc/2KG4-HWDR]. Similarly, YouTube reports that algorithmic changes resulted in U.S. users spending on average 70 percent less time watching "borderline" videos. Greg Bensinger, *YouTube says viewers are spending less time watching conspiracy theory videos. But many still do*, Wash. Post (Dec. 3, 2019, 12:00 PM), https://www.washingtonpost.com/technology/2019/12/03/youtube-says-viewers-are-spending-less-time-watching-conspiracy-videos-many-still-do/ [https://perma.cc/LF4C-MJ2U].

16 Renée DiResta, *Free Speech Is Not the Same As Free Reach*, Wired (Aug. 30, 2018, 4:00 PM), https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/ [https://perma.cc/63EV-KMV7].

17 47 U.S.C. § 230.

18 David McCabe, *Tech Companies Shift Their Posture on a Legal Shield, Wary of Being Left Behind*, N.Y. Times (Dec. 15, 2020), https://www.nytimes.com/2020/12/15/technology/tech-section-230-congress.html [https://perma.cc/FM86-NKYG].

19 Daphne Keller, *One Law, Six Hurdles: Congress's First Attempt to Regulate Speech Amplification in PADAA*, Ctr. Internet Soc'y Blog (Feb. 1, 2021, 5:00 AM), http://cyberlaw.stanford.edu/blog/2021/02/one-law-six-hurdles-congresss-first-attempt-regulate-speech-amplification-padaa [https://perma.cc/XQ4Q-988J].

20 *Contra* Jennifer Cobbe & Jatinder Singh, *Regulating Recommending: Motivations, Considerations, and Principles*, 10 Eur. J. L. Tech. (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3371830 [https://perma.cc/N8SJ-EEB7] (suggesting that regulation of amplification algorithms "can largely sidestep these freedom of expression problems and focus on the use of technical systems by private corporations to pursue their own business goals").

21 529 U.S. 803, 812 (2000); *see also* Sorrell v. IMS Health Inc., 564 U.S. 552, 566 (2011) (stating that government "may no more silence unwanted speech by burdening its utterance than by censoring its content"); Lamont v. Postmaster General, 381 U.S. 301, 309 (1965) (Brennan, J., concurring) (striking down limitation on post office delivery of communist materials and rejecting government's argument that "only inconvenience and not an abridgment is involved").

22 City of Cincinnati v. Discovery Network, Inc., 507 U.S. 410, 418 (1993) (applying heightened scrutiny to "a categorical prohibition on the use of newsracks to disseminate commercial messages").

23 Simon & Schuster, Inc. v. Members of N.Y. State Crime Victims Bd., 502 U.S. 105, 115-16 (1991) (striking down New York's "Son of Sam" law and noting that "[a] statute is presumptively inconsistent with the First Amendment if it imposes a financial burden on speakers because of the content of their speech").

24 Pursuing America's Greatness v. Fed. Election Comm'n, 831 F.3d 500, 510-512 (D.C. Cir. 2016).

25 *Id.* at 510; *cf.* Matal v. Tam, 137 S. Ct. 1744, 1754 (2017) (striking down rule excluding lawful but indecent terms from use as federally protected trademarks).

26 Knight First Amendment Inst. v. Trump, 928 F.3d

226, 238-39 (2d. Cir. 2019).

27 Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017) (noting lower court's conclusion that "the law leaves open adequate alternative means of communication because it permits petitioner to gain access to websites that the court believed perform the 'same or similar' functions as social media").

28 Daphne Keller, *Internet Platforms: Observations on Speech, Danger, and Money* (Aegis Series Paper No. 1807, 2018), https://www.hoover.org/research/internet-platforms-observations-speech-danger-and-money [https://perma.cc/98JT-BETK]; Daphne Keller, *Who Do You Sue?* (Aegis Series Paper No. 1902, 2019), https://www.hoover.org/research/who-do-you-sue [https://perma.cc/46GN-6UZE]; Joris van Hoboken & Daphne Keller, *Design Principles for Intermediary Liability* (Transatlantic Working Grp., Working Paper, Oct. 8, 2019), https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/Intermediary_Liability_TWG_van_Hoboken_Oct_2019.pdf [https://perma.cc/2GWR-QHF6]; Daphne Keller, *Build Your Own Intermediary Law*, Balkinization (June 11, 2019, 1:30 PM), https://balkin.blogspot.com/2019/06/build-your-own-intermediary-liability.html [https://perma.cc/VL6P-FMW4].

29 Daphne Keller, *Empirical Evidence of Over-Removal By Internet Companies Under Intermediary Liability Laws*, Ctr Internet Soc'y Blog (Feb. 8, 2021, 5:11 AM), http://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws [https://perma.cc/SMM8-SDQ2]. These over-removal patterns would presumably be even more common under looser standards, for example assigning liability or removing immunity based on "recklessness" or lack of "reasonableness."

30 José Miguel Vivanco, *Censorship in Ecuador has made it to the Internet*, Hum. Rts. Watch (Dec. 15, 2014, 10:56 AM), https://www.hrw.org/news/2014/12/15/censorship-ecuador-has-made-it-internet [https://perma.cc/65B6-J9PM].

31 Andrea Fuller et al., *Google Hides News, Tricked by Fake Claims*, Wall St. J. (May 15, 2020, 11:43 AM), https://www.wsj.com/articles/google-dmca-copyright-claims-takedown-online-reputation-11589557001 [https://perma.cc/ZGY8-M52G].

32 *See e.g.* Magyar Tartalomszolgáltatók Egyesülete (MTE) v. Hungary, App. No. 22947/13, Eur. Ct. H.R. ¶ 135 (2016), http://hudoc.echr.coe.int/eng?i=001-167828 [https://perma.cc/AF66-8QPN] (rejecting strict liability and de facto monitoring requirement in defamation case as inconsistent with users' human rights); *see also* Case C-360/10 Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV, [2012] 2 C.M.L.R. 18 ¶ 50 (monitoring requirements burden users' freedom of information). *But see* Delfi AS v. Estonia, App. No. 64569/09, Eur. Ct. H.R. ¶ 115 (accepting strict liability and monitoring requirement in case involving hate speech) (2015), http://hudoc.echr.coe.int/webservices/content/pdf/001-155105 [https://perma.cc/6AVY2YHX]; Case C-18/18 Eva Glawischnig-Piesczek v Facebook Ireland Limited EU:C:2019:821 (holding monitoring requirement permissible without discussing user rights).

33 Smith v. California, 361 U.S.147, 153-54 (1959); *see also* Bantam Books v. Sullivan, 371 U.S. 58 (1963); Center for Democracy & Tech. v. Pappert, 337 F.Supp. 2d 606 (E.D. Pa. 2004).

34 Midwest Video Corp. v. FCC, 571 F.2d 1025, 1056-57 (8th Cir. 1979), *aff'd*, 440 U.S. 689 (1979) (rejecting both common carriage obligations for cable companies and imposition of liability for obscene and other unlawful content on public access channels) (resolved on other grounds by Sup Ct.). The 8th Circuit ultimately ruled only on statutory grounds, despite asserting that, "[w]ere it necessary to decide the issue, the present record would render the intrusion represented by the present rules constitutionally impermissible." *Id.* at 1056.

35 *Manila Principles*, ManilaPrinciples.Org, https://www.manilaprinciples.org/ [https://perma.cc/B3HL-5DU9] (last visited May 6, 2020); David Kaye (Special Rapporteur), Hum. Rts. Council, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression* at 6, U.N. Doc. A/HRC/32/38 (May 11, 2016), http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/32/38 [https://perma.cc/G2GH-KAGV] (explaining that the Manila Principles "establish baseline protection for intermediaries in accordance with freedom of expression standards").

36 Llansó et al., *supra* note 7, at 3 ("Content regulation via ranking decisions is particularly prob-

lematic due to the lack of transparency."); *id.* at 20 ("Downranking also raises many of the same free speech issues as content moderation: it prevents platform users from effectively making their voices heard, with little to no accountability when their content is removed.").

37 *See e.g.* Corte Suprema de Justicia de la Nación [CSJN] [National Supreme Court of Justice], 28/10/2014, "Rodríguez, María Belén c. Google Inc. / daños y perjuicios," http://www.saij.gob.ar/corte-supremajusticia-nacion-federal-ciudad-autonoma-buenos-aires-rodriguez-maria-belen-google-incotro-danos-perjuicios-fa14000161-2014-10-28/123456789-161-0004-1ots-eupmocsollaf [https://perma.cc/6876-2G3P] (Arg.) (to protect user rights, platforms must not be obligated to take down most information unless a court has deemed it illegal); *Belen Rodriguez – English Translation*, Stan. Ctr. for Internet and Soc'y (2017), https://docs.google.com/document/d/10BlZTeIPXqPGWvYUFzNgjDMZaPOx9bneAhK4f3FCLkg/edit; Singhal v. Union of India, (2015) 12 SCC 73, ¶¶ 100, 117 (India) (requiring review by court or appropriate public authority); Law No. 20.435 Art. 85 (Chile) (judicial review for copyright takedown requests); Marco Civil da Internet, Federal Law no. 12.965 (Braz.) (judicial review for takedown requests other than child sexual abuse material, non-consensual sexual images, and copyright); *Marco Civil English Version*, Pub. Knowledge (May 27, 2014), https://www.publicknowledge.org/documents/marco-civil-english-version/ [https://perma.cc/6HMS-JS5Y].

38 *See, e.g.*, United States v. Stevens, 559 U.S. 460, 470 (2010) (rejecting "a categorical balancing of the value of the speech against its societal costs" as a "startling and dangerous" approach to First Amendment law). The U.S. approach comes with costs—even from a pure pro-expression perspective. As one dissenting U.S. Supreme Court Justice noted, by eschewing a "middle way" between the extremes of "ban totally or do nothing at all," the Court potentially encourages more enforcement against speech that might otherwise have been rendered less available, and thus more tolerated. Ashcroft v. Am. Civ. L. Union, 542 U.S. 656, 691 (2004) (Breyer, J., dissenting) (striking down law requiring online pornography sites to verify users' ages).

39 United States v. Playboy, 529 U.S. 803, 827 (2000). If we consider speech presented to a captive audience to be a form of amplification, then one possible exception is Lehman v. Shaker Heights, 418 U.S. 298 (1974), which upheld a prohibition on political ads on buses.

40 *Playboy*, 529 U.S. at 804 (noting the "key difference between cable television and the broadcasting media, which is the point on which this case turns: Cable systems have the capacity to block unwanted channels on a household-by-household basis"); Ashcroft v. Am. Civ. L. Union, 542 U.S. 656, 667-70 (2004). If other laws establish the possibility of less restrictive responses to similar problems—as the DMCA arguably does for any changes to CDA 230—then that, too, is relevant to courts' scrutiny. Denver Area Educ. Telecomm. Consortium, Inc. v. F.C.C., 518 U.S. 727, 756 (1996) (holding that the law requiring cable companies to segregate offensive content on leased access channels violates the First Amendment).

41 evelyn douek, *Governing Online Speech: From "Posts as Trumps" to Proportionality and Probability*, 121 Colum. L. Rev. (forthcoming 2021) (discussing probabilistic nature and error acceptance in algorithmic design).

42 Human Rights Watch, *"Video Unavailable": Social Media Platforms Remove Evidence of War Crimes* (Hum. Rts. Watch 2020), https://www.hrw.org/sites/default/files/media_2020/09/crisis_conflict0920_web_0.pdf [https://perma.cc/9P34-FH9C].

43 Maarten Sap et al., *The Risk of Racial Bias in Hate Speech Detection*, Proc. 57th Ann. Meeting Ass'n for Computational Linguistics 1668 (July 2019), https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf [https://perma.cc/2238-T2ME].

44 *See* Letter from David Kaye, Special Rapporteur Promotion and Prot. Right to Freedom Op. and Expression, U.N., et al., to Off. of the High Comm'r, U.N. Hum. Rts. (Dec. 7, 2018) (https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=2423 [https://perma.cc/D85A-2A4W]); *Joint Letter to EU Parliament: Vote Against Proposed Terrorist Content Online Regulation*, Hum. Rts. Watch (Mar. 25, 2021, 1:00 AM), https://www.hrw.org/news/2021/03/25/joint-letter-eu-parliament-vote-against-proposed-terrorist-content-online [https://perma.cc/EJU3-KA95]. *Compare*

Tech Against Terrorism, *Content personalisation and the online dissemination of terrorist and violent extremist content* 1-2 (Tech Against Terrorism Position Paper 2021), https://www.techagainstterrorism.org/2021/02/17/position-paper-content-personalisation-and-the-online-dissemination-of-terrorist-and-violent-extremist-content/ [https://perma.cc/VV4S-MPJD] (reviewing empirical research and finding "very limited evidence" to support focus on personalization and amplification as drivers of violent extremism), *with* Michal Lavi, *Do Platforms Kill?*, 43 Harv. J. L. Pub. Pol'y 477 (2020) (asserting that platforms' algorithms have the ability to "promote efficient identification and removal of terrorist content" and proposing legal requirement to "disable the ability to create unlawful recommendations").

45 Samantha Cole, *Craigslist Just Nuked Its Personal Ads Section Because of a Sex-Trafficking Bill*, Vice (Mar. 23, 2018, 5:18 AM), https://www.vice.com/en/article/wj75ab/craigslist-personal-ads-sesta-fosta [https://perma.cc/H487-SLUW]; Makena Kelly, *Democrats want data on how sex workers were hurt by online crackdown*, Verge (Dec. 17, 2019, 4:12 PM), https://www.theverge.com/2019/12/17/21026787/sesta-fosta-congress-study-hhs-sex-work-ro-khanna-elizabeth-warren-ron-wyden [https://perma.cc/2AZH-WNCD].

46 Press Release, Knight Found., Americans Support Free Speech Online but Want More Action to Curb Harmful Content (June 16, 2020), https://knightfoundation.org/press/releases/americans-support-free-speech-online-but-want-more-action-to-curb-harmful-content/ [https://perma.cc/GYX9-FEE2].

47 Joined Cases C-236/08C-238/08, Google France SARL v. Louis Vuitton Malletier SA, 2010 E.C.R. I-2417; The Digital Services Act package, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package [https://perma.cc/Z2HU-LAUD] (last visited May 26, 2021).

48 *See e.g.*, Zhang v. Baidu, 10 F.Supp. 3d 433 (S.D.N.Y 2014); e-ventures Worldwide v. Google, Inc., No. 2:14-cv-646-FtM-PAM-CM, 2017 WL 2210029 (M.D. Fla. Feb. 8, 2017); Keller, *Who Do You Sue?*, *supra* note 28 at 17-22 (discussing case law).

49 Turner Broadcasting Sys., Inc. v. F.C.C., 512 U.S. 622 (1994); Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., 515 U.S. 557 (1995).

50 U.S. Copyright Off., Circular 14, *Copyright in Derivative Works and Compilations* (July 2020), https://www.copyright.gov/circs/circ14.pdf [https://perma.cc/2M77-DX64].

51 Dep't for Digit., Culture, Media and Sport, *Online Harms White Paper: Full government response to the consultation* (Dec. 15, 2020) (unpublished white paper), https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response.

52 Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos, ECLI:EU:C:2014:317. *See* Daphne Keller, *The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 Berkeley Tech. L. J. 287, 312-17 (2018); Robert C. Post, *Data Privacy and Dignitary Privacy: Google Spain, the Right To Be Forgotten, and the Construction of the Public Sphere*, 67 Duke L. J. 981 (2018).

53 *Google Spain*, ECLI:EU:C:2014:317 at ¶ 88. *Cf.* Wegrzynowski & Smolczewski v. Poland, App. No. 33846/07, Eur. Ct. H.R. (2013), http://hudoc.echr.coe.int/eng?i=001-122365 [https://perma.cc/2GQV-DEER] (Under European human rights law, plaintiffs may obtain annotation, but not deletion, of defamatory articles in a news archive.). *See also* Ellen Goodman, *Digital Information Fidelity and Friction*, Knight First Amendment Inst. at Colum. Univ. (Feb. 26, 2020) https://knightcolumbia.org/content/digital-fidelity-and-friction [https://perma.cc/245X-8XQM] (discussing friction in analog and digital information distribution systems).

54 Theo Bertram et al., *Three Years of the Right to be Forgotten* (Google unpublished working paper, 2018), https://d110erj1750600.cloudfront.net/wp-content/uploads/2018/02/google.pdf [https://perma.cc/EZ6E-MUML].

55 Council Directive 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, 2018 O.J. (L 303/69).

56 Will Goodbody, *Aspects of online safety*

bill 'vague, open-ended' - IHREC, RTE (Mar. 15, 2021, 7:12 AM), https://www.rte.ie/news/ireland/2021/0315/1204012-online-safety/ [https://perma.cc/5FB7-CAJF].

57 Requests to delist content under European privacy law, https://transparencyreport.google.com/eu-privacy/overview?hl=en [https://perma.cc/L2NV-WWQG] (last visited May 6, 2020).

58 Niall McCarthy, *Facebook Removes Record Number Of Hate Speech Posts [Infographic]*, Forbes (May 13, 2020, 6:59 AM), https://www.forbes.com/sites/niallmccarthy/2020/05/13/facebook-removes-record-number-of-hate-speech-posts-infographic/?sh=5b7027083035 [https://perma.cc/WUJ7-2K4T].

59 Lyrissa Barnett Lidsky, *Incendiary Speech and Social Media*, 44 Tex. Tech. L. Rev. 147, 148 (2011).

60 Dennis v. United States, 341 U.S. 494, 579-581 (1978) (Black, J., dissenting) (noting legality of books by Marx, Engels, Lenin, and Stalin).

61 F.C.C. v. Pacifica Found., 438 U.S. 726 (1978); Sable Commc'ns of Cal., Inc. v. F.C.C., 492 U.S. 115, 128 (1989) (distinguishing potentially "invasive" nature of radio from adult telephone services sought out by willing listeners). Jack Balkin breaks down the different potential meanings of this "pervasiveness" justification in *Media Filters, the V-Chip, and the Foundations of Broadcast Regulation*, 45 Duke L. J. 1131 (1996).

62 Red Lion Broadcasting Co. v. F.C.C., 395 U.S. 367 (1969); Turner Broadcasting Sys., Inc. v. F.C.C., 520 U.S. 180 (1997); Keller, *Who Do You Sue?*, *supra* note 28. Proponents of laws restricting amplification of otherwise lawful speech could also present them as a form of "zoning," justified by secondary effects like the decline in neighborhood property values recognized as a basis for zoning restrictions on adult theaters in City of Renton v. Playtime Theatres, Inc., 475 U.S. 41 (1986). I don't think this would work: The secondary effects of online hate speech, disinformation, and the like are either (1) themselves speech-related, like "political polarization" or "decline of public discourse" or (2) already addressed in other areas of First Amendment law, like incitement to violence. But others may wish to explore this avenue in more depth.

63 Turner Broadcasting Sys., Inc. v. F.C.C., 512 U.S. 622, 632, 633 (1994); *Turner*, 520 U.S. at 191. *Cf.* United States Telecomm. Ass'n v. F.C.C., 825 F.3d 674 (2016) (en banc) (rejecting internet service provider's First Amendment objection to net neutrality, which prevented both excluding and "throttling" or slowing disfavored content).

64 Reno v. American Civ. L. Union, 521 U.S. 844, 868 (1997).

65 *Id.* at 867.

66 *Cf.* John Blevins, *The New Scarcity: A First Amendment Framework for Regulating Access to Digital Media Platforms*, 79 Tenn. L. Rev. 353, 397 (2012) (describing similar regulations of search algorithms as "impossible to design and administer").

67 Facebook: Turn Off The Algorithms, https://accountabletech.org/campaign/facebook-turn-off-the-algorithms/ [https://perma.cc/Q299-S78T] (last visited May 6, 2020).

68 Protecting Americans from Dangerous Algorithms Act, H.R. 8636, 116th Cong. (2020) (immunizing chronological but not other algorithmic ranking).

69 As Benedict Evans notes, for users with high-volume feeds a chronological presentation would be "not so much chronological in any useful sense as a random sample, where the randomiser is simply whatever time you yourself happen to open the app[.]" Benedict Evans, *Death of the Newsfeed*, Benedict Evans (Apr. 3, 2018), https://www.ben-evans.com/benedictevans/2018/4/2/the-death-of-the-newsfeed [https://perma.cc/B54F-GUVN].

70 Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, Facebook (May 5, 2021, 5:38 PM), https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/.

71 *Cf.* Balkin, *supra* note 61, at 1171 (discussing filmmakers' incentives to come close to the line of content restricted by MPAA guidelines).

72 Knight First Amend. Inst. at Colum. Univ. v. Trump, 953 F.3d 216, 217 (2d Cir. 2020).

73 *Transparency Report 2019* (Reddit May 6, 2021), https://www.redditinc.com/policies/transparency-report-2019 [https://perma.cc/M9GA-4MFR].

74 Noam Cohen, *Google Halts 'Miserable Failure' Link to President Bush*, N.Y. Times (Jan. 29, 2007), https://www.nytimes.com/2007/01/29/technology/

29google.html [https://perma.cc/F25C-8WRG].

75 *See* Llansó et al., *supra* note 7, at 16; Leerssen, *supra* note 7, at 5-6.

76 evelyn douek, *What Does "Coordinated Inauthentic Behavior" Actually Mean?*, SLATE (July 2, 2020, 5:26 PM), https://slate.com/technology/2020/07/coordinated-inauthentic-behavior-facebook-twitter.html [https://perma.cc/3UU8-EFV2].

77 Prager Univ. v. Google LLC, 951 F.3d 991, 998 (9th Cir. 2020).

78 *See, e.g.*, GOOGLE, SEARCH QUALITY EVALUATOR GUIDELINES 134 (Dec. 2019) (assuming "dominant educational/informational intent" as ranking priority when users appear to be seeking hateful or intolerant content).

79 *Cf.* Leathers v. Medlock, 499 U.S. 439, 448 (1991) (taxing cable television without similarly taxing other media "does not implicate the First Amendment unless it discriminates on the basis of ideas[,]" but noting there was "no indication in this case that [lawmakers] targeted cable television in a purposeful attempt to interfere with its First Amendment activities").

80 *See, e.g.*, Ward v. Rock Against Racism, 491 U.S. 781, 792 (1989); Barr v. Am. Ass'n of Pol. Consultants, Inc., 140 S. Ct. 2335, 2346 (2020) ("[A] law banning the use of sound trucks for political speech—and only political speech—would be a content-based regulation, even if it imposed no limits on the political viewpoints that could be expressed.") (quoting Reed v. Gilbert, 576 U.S. 155, 168 (2015)). "Volume," like "amplification," can have subtly different meanings. *See* Miller, *supra* note 7, at 13 (noting Supreme Court's divergent analysis in cases involving sonic amplification in public forums on one hand, and numerical amplification of audience size via campaign spending and other paid uses of privately owned media on the other).

81 17 U.S.C. §§ 1001-10. Copyright law, which restricts duplication, may generally be a fruitful point of comparison for potential amplification laws.

82 Kovacs v. Cooper, 336 U.S. 77, 102-103 (1949) (Black, J., dissenting). The Court more recently declined to resolve a state's argument that a law restricting sex offenders' access to social media was, like laws against sound trucks and amplifiers, a neutral "time, place, and manner" restriction. Packingham v. North Carolina, 137 S. Ct. 1730, 1739 (2017) (Alito, J., concurring) (noting state's argument but concluding, with the majority of the court, that it need not be addressed because the law at issue failed even intermediate scrutiny).

83 *See* Ranking Digit. Rts., *It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy* (Ranking Digit. Rts. 2020), https://rankingdigitalrights.org/its-the-business-model/ [https://perma.cc/ZU4X-EYS3]; Jeff Gary & Ashkan Soltani, *First Things First: Online Advertising Practices and Their Effects on Platform Speech*, KNIGHT FIRST AMENDMENT INST. AT COLUM. UNIV. (Aug. 21, 2019), https://knightcolumbia.org/content/first-things-first-online-advertising-practices-and-their-effects-on-platform-speech [https://perma.cc/9H7R-F5MQ].

84 *Draft Report*, *supra* note 12, at 5.

85 Ben Popken, *As algorithms take over, YouTube's recommendations highlight a human problem*, NBC NEWS (Apr. 19, 2018, 3:14 PM), https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n867596 [https://perma.cc/5XU3-5KJT].

86 Complicating factors for ad-driven businesses include long-term customer retention as well as many advertisers' demands for platforms to remove content such as extremism or hate speech. Olivia Solon, *Google's bad week: YouTube loses millions as advertising row reaches US*, GUARDIAN (Mar. 25, 2017, 6:00 AM), https://www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon [https://perma.cc/6Q46-HZAM]; Alex Hern, *Third of advertisers may boycott Facebook in hate speech revolt*, GUARDIAN (June 30, 2020, 11:15 AM), https://www.theguardian.com/technology/2020/jun/30/third-of-advertisers-may-boycott-facebook-in-hate-speech-revolt [https://perma.cc/V6T2-KGGP].

87 Samuel Stolton, *'No longer acceptable' for platforms to take key decisions alone, EU Commission says*, EURACTIV (Jan. 20, 2021), https://www.euractiv.com/section/digital/news/no-longer-acceptable-for-platforms-to-take-key-decisions-alone-eu-commission-says/ [https://perma.cc/JME5-LSP6].

88 *See generally* Robert Sopolsky, Behave: The Biology of Humans at Our Best and Worst (2017).

89 Sander van der Linden, *How the Illusion of Being Observed Can Make You a Better Person*, Sci. Am. (May 3, 2011), https://www.scientificamerican.com/article/how-the-illusion-of-being-observed-can-make-you-better-person/ [https://perma.cc/9NKH-X8ZB]; Keith Dear et al., *Do 'watching eyes' influence antisocial behavior? A systematic review and meta-analysis*, 40 Evolution Hum. Behav. 269 (2019). This might explain why 13% of searches we conduct from our browsers are for porn—and why that number rises to 20% when we search from the relative seclusion of our mobile devices. Katharina Buchholz, *How Much of the Internet Consists of Porn?*, Statista (Feb. 11, 2019), https://www.statista.com/chart/16959/share-of-the-internet-that-is-porn [https://perma.cc/9JCR-EBSR].

90 James Vincent, *Twitter is bringing its 'read before you retweet' prompt to all users*, Verge (Sept. 25, 2020, 7:08 AM), https://www.theverge.com/2020/9/25/21455635/twitter-read-before-you-tweet-article-prompt-rolling-out-globally-soon [https://perma.cc/PXD3-TZ4Q].

91 Nick Clegg, *You and the Algorithm: It Takes Two to Tango*, Medium (Mar. 31, 2021), https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2 [https://perma.cc/G3HM-EU2S].

92 The Digital Services Act package, Art. 36, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package [https://perma.cc/777Q-RHTW] (last visited May 26, 2021).

93 Reno v. American Civ. L. Union, 521 U.S. 844, 868 (1997); United States v. Playboy, 529 U.S. 803, 815 (2000); Ashcroft v. Am. Civ. L. Union, 542 U.S. 656, 667-70 (2004).

94 Facebook reports that one of its voluntary changes, demoting content from followed pages including news sources, decreased user time on the platform by 50 million hours each day and cut billions of dollars from its market capitalization. Clegg, *supra* note 91. *See also* Shana Lebowitz, *A former Google data scientist explains why Netflix knows you better than you know yourself*, Bus. Insider India (June 2, 2018), https://www.businessinsider.in/A-former-Google-data-scientist-explains-why-Netflix-knows-you-better-than-you-know-yourself/articleshow/58964661.cms [https://perma.cc/4RMU-3G7M] (noting divergence between people's expressed preferences and the revealed preferences suggested by their behavior).

95 There would remain a risk of moderation decisions flowing from better-resourced platforms to their smaller counterparts. *See generally* evelyn douek, *The Rise of Content Cartels*, Knight First Amendment Inst. at Colum. Univ. (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels [https://perma.cc/N2M8-XAT9].

96 Associated Press v. United States, 326 U.S. 1 (1945).

97 Mike Masnick, *Protocols, Not Platforms: A Technological Approach to Free Speech*, Knight First Amendment Inst. at Colum. Univ. (Aug. 19, 2020), https://knightcolumbia.org/content/protocols-not-platforms-a-technological-approach-to-free-speech [https://perma.cc/GAK4-4QLF]; Daphne Keller, *If Lawmakers Don't Like Platforms' Speech Rules, Here's What They Can Do About It*, Techdirt (Sept. 9, 2020, 12:00 PM), https://www.techdirt.com/articles/20200901/13524045226/if-lawmakers-dont-like-platforms-speech-rules-heres-what-they-can-do-about-it-spoiler-options-arent-great.shtml [https://perma.cc/29KK-JEKC]; *Testimony of Jack Dorsey, Chief Executive Officer, Twitter, Inc. Before the U.S. S. Judiciary Comm.*, 116th Cong. (Nov. 17, 2020), https://www.judiciary.senate.gov/imo/media/doc/Dorsey%20Testimony.pdf [https://perma.cc/C7TV-DBA3]; Francis Fukuyama, *Making the Internet Safe for Democracy*, 32 J. Democracy 37 (2021).

98 *See* Keller, *supra* note 97.

99 Stephen Wolfram, T*estifying at the Senate about A.I.*, Stephen Wolfram Writings (June 25, 2019), https://writings.stephenwolfram.com/2019/06/testifying-at-the-senate-about-a-i-selected-content-on-the-internet [https://perma.cc/PR82-38LA].

100 Lucas Matney, *Twitter's decentralized future*, TechCrunch (Jan. 15, 2021, 2:52 PM), https://techcrunch.com/2021/01/15/twitters-vision-of-decentralization-could-also-be-the-far-rights-internet-endgame/ [https://perma.cc/YTR9-FUC3].

101 *See, e.g.*, Ian Brown, *Interoperability as a tool for competition regulation* at 33-39, Openforum Acade-

my (2020), https://osf.io/preprints/lawarxiv/fbvxd/ [https://perma.cc/3DXL-8TR2].

102 *See* Andrew Tutt, *An FDA for Algorithms*, 69 Admin. L. Rev. 83 (2017).

103 Jeff Horwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, Wall St. J. (May 26, 2020, 11:38 AM) https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499 [https://perma.cc/K6XQ-GN93] (discussing Facebook's internal use of "eat your veggies" term).

104 Smith v. California, 361 U.S. 147, 152-53 (1959); *cf.* Tutt, *supra* note 102.

105 An "information fiduciaries" model, which creates duties toward platform users, might also be characterized as a consumer protection measure. *Jonathan Zittrain and Jack Balkin Propose Information Fiduciaries to Protect Individual Privacy Rights*, Tech. Acad. Pol'y (Sept. 28, 2018), https://www.techpolicy.com/Blog/September-2018/Jonathan-Zittrain-and-Jack-Balkin-Propose-Informat.aspx [https://perma.cc/GSR7-Y8J3]; *cf.* James Grimmelmann, *Speech Engines*, 98 Minn. L. Rev. 868 (2014). But it does not create a back door for government to limit lawful speech. In any case, a platform that shielded a user from lawful information that she appeared to want would be less like a fiduciary, and more like a parent or caretaker.

106 Wash. Post v. McManus, 944 F.3d 506 (4th Cir. 2019) (striking Maryland's campaign advertising transparency law); *see generally* Valerie C. Brannon, Cong. Rsch. Serv., R45700, Assessing Commercial Disclosure Requirements Under the First Amendment (2019), https://crsreports.congress.gov/product/pdf/R/R45700 [https://perma.cc/RBB6-MZV5].

107 Daphne Keller, *Six Constitutional Hurdles for Platform Speech Regulation*, Ctr. Internet Soc'y Blog (Jan. 22, 2021, 6:50 AM), http://cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation-0 [https://perma.cc/UQ72-YQMT].

108 Matal v. Tam, 137 S. Ct. 1744, 1760 (2017).

About the Author

Daphne Keller directs the Program on Platform Regulation at Stanford's Cyber Policy Center, and was formerly the director of intermediary liability at the Center for Internet and Society at Stanford Law School. Her work focuses on platform regulation and Internet users' rights. She has published both academically and in popular press; testified and participated in legislative processes; and taught and lectured extensively. Her recent work focuses on legal protections for users' free expression rights when state and private power intersect, particularly through platforms' enforcement of terms of service or use of algorithmic ranking and recommendations. Until 2015 Daphne was associate general counsel for Google, where she had primary responsibility for the company's search products. She worked on groundbreaking intermediary liability litigation and legislation around the world and counseled both overall product development and individual content takedown decisions.

Acknowledgments

Many thanks for feedback from Jack Balkin, Joan Barata, Katy Glenn Bass, Brian Downing, Al Gidari, Eric Goldman, Jameel Jaffer, Dina Lamdany, Paddy Leerssen, Max Levy, Alexander MacGillivray, Nate Persily, Monroe Price, Blake Reid, Thomas Streinz, Eugene Volokh, Nicole Wong, and participants in NYU's Guarini Colloquium and the Yale Information Society Project's Ideas Lunch.

© 2021, Daphne Keller.

About the Knight First Amendment Institute

The Knight First Amendment Institute at Columbia University defends the freedoms of speech and the press in the digital age through strategic litigation, research, and public education. It promotes a system of free expression that is open and inclusive, that broadens and elevates public discourse, and that fosters creativity, accountability, and effective self-government.

knightcolumbia.org

Design: Point Five
Illustration: © Dale Crosby-Close


KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY