5

Social Media as the New Gatekeepers

As discussed in the previous chapter, the preferred regulatory solution to the supposed ills of social media proposed by conservative critics is to strip social media platforms of almost all editorial power. Their progressive counterparts (such as Senator Elizabeth Warren of Massachusetts and Senator Amy Klobuchar of Minnesota), perhaps unsurprisingly, propose exactly the opposite. They want platforms to engage in *more* content moderation, under the guidance of experts and political leaders (such as, presumably, themselves). In other words, progressives want platform owners to operate as gatekeepers of speech and knowledge,[1] excluding from public discourse the kinds of harmful content, described in Chapter 2, to which they object. As this chapter will demonstrate, however, the progressive proposals make as little sense as the conservative ones.

5.1 THE GOVERNMENT AS GATEKEEPER

At the outset, it is important to draw an important distinction between voluntary actions taken by platforms to moderate content (which, for reasons described in Chapter 4, they have a constitutional right to take) and government mandates requiring platforms to block or deemphasize specific, harmful content. While most progressive calls for content moderation seem to envision voluntary steps by platforms, critics sometimes veer into legislative mandates. For example, in 2021 in the midst of the COVID-19 pandemic, Senator Klobuchar introduced legislation that would have made platforms liable for any health misinformation that they algorithmically promote – specifically,

[1] By gatekeepers, I mean entities and/or institutions who control what information and what sources of information the general public is exposed to without great effort on the audience's part.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

the legislation would have stripped platforms of the immunity they normally enjoy under Section 230 of the Communications Decency Act regarding such information.[2]

Similarly, in 2020 a group of twenty state Attorneys General sent a joint letter to Facebook, calling on the firm to make greater efforts to block allegedly harmful content.[3] The letter identified a number of different kinds of such harmful speech, including disinformation, cyberstalking, doxing (publishing private information), and swatting (filing false police reports), but the primary focus of the letter was hate speech – which is to say, speech that vilifies specific groups based on characteristics such as race, national origin, sex, religion, or sexual orientation.[4] And while the letter itself did not go beyond calling on Facebook to take voluntary action, in an interview with the New York Times, then-Attorney General Gurbir S. Grewal of New Jersey (one of the signatories) threatened that, if Facebook did not act, "we always have a variety of legal tools at our disposal."[5] In other words, Grewal appeared to be suggesting that, if Facebook failed to do a better job of blocking hate speech and other harmful content, state prosecutors would seek legal remedies against it, thus opening the door to direct legal requirements to engage in content moderation.

The problem with proposals such as these, however, is that they are quite clearly in violation of the First Amendment. It is textbook law that aside from a very few, narrowly defined categories of speech such as obscenity,[6] child pornography,[7] and threats of violence,[8] the First Amendment protects all speech from government censorship. The Supreme Court has made it clear, moreover, that disinformation – which is to say intentional lies – are *not* an unprotected category, and so enjoy full First Amendment protections.[9] Furthermore, despite common misunderstandings among the public and

[2] *Health Misinformation Act of 2021*, S. 2448, 117th Congress (July 22, 2021), www.congress.gov/bill/117th-congress/senate-bill/2448/text.

[3] Davey Alba, *Facebook Must Better Police Online Hate, State Attorneys General Say*, N.Y. Times (Aug. 5, 2020), www.nytimes.com/2020/08/05/technology/facebook-online-hate.html.

[4] Letter from Karl A. Racine, Attorney General, District of Columbia, Kwame Raoul, Attorney General, State of Illinois, Gurbir S. Grewal, Attorney General, State of New Jersey, et al., to Mark Zuckerberg, Chairman and Chief Executive Officer, Sheryl Sandberg, Chief Operating Officer (Aug. 5, 2020), https://int.nyt.com/data/documenttools/facebook-attorneys-general-letter/5073887056 2dec84/full.pdf.

[5] Alba, *supra* n. 3.

[6] Roth v. United States, 354 U.S. 476 (1957).

[7] Ferber v. New York, 458 U.S. 747 (1982).

[8] Counterman v. Colorado, 600 U.S. 66 (2023).

[9] United States v. Alvarez, 567 U.S. 709 (2012).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

even elected leaders and some lawyers,[10] the Supreme Court has also made it clear, unanimously no less, that "hate speech" is also fully protected by the First Amendment in the United States (unlike in many other countries).[11] Indeed, it is not entirely clear that even doxing is unprotected speech, though on this point the law is somewhat unclear.

The final piece of the puzzle is that under long-standing First Amendment doctrine, when the government singles out protected speech for regulation based on the content or especially the viewpoint of the speech, the regulatory effort is almost always invalidated by courts.[12] Yet by singling out health disinformation, Senator Klobuchar's legislative proposal clearly singles out specific content. And as for former Attorney General Grewal's threats regarding hate speech, current law treats regulations of hate speech as viewpoint-based, and so essentially automatically unconstitutional.[13]

The implications of all of this are clear: Just as the government could not directly prohibit or restrict "disinformation" or "hate speech," it also cannot require social media platforms to remove or limit the availability of such speech. To conclude otherwise would be to permit the government to involuntarily dragoon private companies to do things that the Constitution forbids the government. But just as the government cannot require private security firms to conduct searches that violate the Fourth Amendment, so it cannot require platforms to engage in censorship that violates the First. Then-Attorney General Grewal's threats in that regard were, then, just so much hot air.

Senator Klobuchar's proposal, which could not have completely banned disinformation but merely removed it from the Section 230 protective umbrella, poses a slightly more difficult question but ultimately meets the same fate (the details of how Section 230 operates, as well as proposals to amend it, are the topics of the next chapter). But to begin with, it is not even clear what the legislation would accomplish. In theory, it would open platforms up to liability for spreading health disinformation; but what kind of liability? We do not typically impose civil liability even on *speakers* based on protected speech, or spreaders

[10] Eugene Volokh, *No, Gov. Dean, There Is No "Hate Speech" Exception to the First Amendment*, Washington Post (April 21, 2017), www.washingtonpost.com/news/volokh-conspiracy/wp/2017/04/21/no-gov-dean-there-is-no-hate-speech-exception-to-the-first-amendment/; Eugene Volokh, *California AG's Brief Claims "Hate Speech" Is Constitutionally Unprotected*, Reason.com (Nov. 25, 2020), https://reason.com/volokh/2020/11/25/california-ags-brief-claims-hate-speech-is-constitutionally-unprotected/.

[11] Matal v. Tam, 582 U.S. 218 (2017); *see also* R.A.V. v. City of St. Paul, 505 U.S. 377 (1992).

[12] Reed v. Town of Gilbert, 576 U.S. 155 (2015).

[13] *Matal*, 582 U.S. at 243 (plurality opinion); *ibid.* at 248–49 (Kennedy, J., concurring in part and concurring in the judgment).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

of vaccine disinformation such as Robert F. Kennedy, Jr.[14] could be constantly in the dock. Given that, it is quite unclear under what theory a mere unwitting distributor of such content, such as a platform, would face liability.

But even if liability was a possibility under tort law, what Senator Klobuchar proposed would likely violate the First Amendment. Congress no doubt has the power to entirely repeal Section 230 and so eliminate platforms' statutory protections from liability for third-party content.[15] Furthermore, Congress also probably has the power to selectively deny Section 230 immunity for speech unprotected by the First Amendment – as Congress in fact did with respect to speech promoting sex trafficking.[16] But to permit Congress to selectively strip immunity for *protected* speech – which would effectively force platforms to block speech it was made aware of that even arguably falls within the disfavored category – would be an extraordinarily potent tool of censorship. Surely the First Amendment could not permit Congress to eliminate Section 230 immunity for, say, posts critical of Democratic elected officials, but not Republican ones. But given the porousness of the definition of "health misinformation" – some platforms labeled posts opposed to mask mandates for children, surely a question on which reasonable people could differ given the limited risks children face from COVID, as misinformation – Senator Klobuchar's proposed law seems to be at least analogous.

In short, government mandates that social media platforms moderate constitutionally protected content are quite clearly unconstitutional. As a consequence, despite much talk among politicians and journalists, such laws have not been enacted in the United States. It is noteworthy in this regard that jurisdictions such as Germany, in which hate speech does *not* receive strong protections, *have* adopted and enforced laws requiring platforms to swiftly

[14] Anjali Huynh, *5 Noteworthy Falsehoods Robert F. Kennedy Jr. Has Promoted*, N.Y. Times (July 6, 2023), www.nytimes.com/2023/07/06/us/politics/rfk-conspiracy-theories-fact-check.html.

[15] Which is not to say that such a repeal would leave platforms liable for all content they host – it seems likely that the First Amendment itself would provide some shield to platforms, though the exact contours of that shield have not been determined because Section 230's existence has made the issue moot.

[16] Tom Jackman, *Trump Signs "FOSTA" Bill Targeting Online Sex Trafficking, Enables States and Victims to Pursue Websites*, Washington Post (April 11, 2018), www.washingtonpost.com/news/true-crime/wp/2018/04/11/trump-signs-fosta-bill-targeting-online-sex-trafficking-enables-states-and-victims-to-pursue-websites/. Ironically, the bill appears to have completely failed to achieve its goals and instead had perverse consequences. Melissa Gira Grant, *The Real Story of the Bipartisan Anti-Sex Trafficking Bill that Failed Miserably on Its Own Terms*, New Republic (June 23, 2021), https://newrepublic.com/article/162823/sex-trafficking-sex-work-sesta-fosta.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

remove hate speech.[17] But in the United States, those who wish platforms to play the role of gatekeepers of constitutionally protected content must necessarily rely on the voluntary cooperation of the platforms themselves. And that, it would seem, is the primary goal of progressive critics of social media: to convince/induce/pressure platforms into performing that role and removing harmful content from public discourse.

One final point on this: While US law draws a strong distinction between government mandates to moderate protected content and encouraging platforms to voluntarily do to same, on the ground the line between these two things is not always clear. In particular, evidence has emerged that during the height of the COVID-19 pandemic, the Biden Administration imposed heavy pressure on the major platforms to remove mis- and disinformation about COVID-19, and COVID vaccines. This eventually led to a lawsuit brought by some social media users whose posts were blocked, as well as Republican politicians, claiming that the Administration's efforts had crossed the line into violating the First Amendment. Ultimately, the Supreme Court dismissed the case on technical grounds, concluding in essence that the plaintiffs had failed to show that government lobbying caused any specific content moderation decisions.[18] Moreover, in a letter to Congress Mark Zuckerberg (the CEO of Meta, which owns both Facebook and Instagram), while expressing regret over his platforms' cooperation with the Administration, insisted that all content moderation decisions had ultimately been made by the platforms themselves, not by the government.[19] But the fact remains that at some point, government cajoling and pressure can cross the line into unconstitutional coercion.

5.2 THE OLD GATEKEEPERS

To understand why there are so many voices, primarily (but not exclusively) on the political left, who endorse a gatekeeper role for platforms, one must take a step back and consider the nature and origins of the gatekeeper function. Underlying the desire to resurrect the gatekeeper function is a pervasive fear among progressive critics, discussed in Chapter 2, that the spread of harmful content is systematically harming society, including especially

[17] *Germany Starts Enforcing Hate Speech Law*, BBC (Jan. 1, 2018), www.bbc.com/news/technology-42510868.

[18] Murthy v. Missouri, 144 S. Ct. 1972 (2024).

[19] Gnaneshwar Rajan and Nandita Bose, *Zuckerberg Says Biden Administration Pressured Meta to Censor COVID-19 Content*, Reuters (Aug. 27, 2024), www.reuters.com/technology/zuckerberg-says-biden-administration-pressured-meta-censor-covid-19-content-2024-08-27/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

vulnerable elements of society such as racial and sexual minorities. But not all "harmful" speech elicits equal amounts of concern. Due perhaps to the COVID-19 pandemic and its corrosive effect on politics and public discourse, there is no question that from the perspective of progressive critics, the most dangerous and corrosive form of online speech is mis- and disinformation, especially (though not exclusively) on matters pertaining to health (as demonstrated by Senator Klobuchar's proposed legislation). Why are falsehoods of such particular concern to these critics?

Driving the campaign against misinformation is a basic, existential worry that the spread of online falsehoods is systematically eroding the common base of facts, accepted by the broader public, that is essential to a system of democratic self-governance. Absent such a common base or factual consensus, it is feared, democratic politics will tend to collapse into polarized camps that cannot accept the possibility of electoral defeat (as they arguably have in recent years in the United States). The only way to combat these developments, these critics believe, is to divert the flow of mis- and disinformation from the public ecosphere. And for that to happen, gatekeepers who can identify and favor trusted and trustworthy sources of information are essential.

But therein lies the nub of the problem. If one wishes to restore a common, factual consensus in the public sphere, one must confront the unavoidable question of "Who to trust?" But underlying *that* question is yet another, more foundational one: "Who *decides* who to trust?" Ultimately, of course, each person must decide for themselves who to trust. But for a societal consensus on this question to emerge, some common source of authority is seemingly needed. If there is one lesson that can be drawn from the modern era of social media, it is that robust, public discourse alone cannot be expected to generate an automatic consensus on who can be trusted (or on what are trustworthy facts). The quest for trusted communicators, then, is in truth a quest for authoritative sources of trust – which is to say, a quest for authority. In the internet era, centralized control over information flows has fragmented and, consequently, so too has the authority to identify trusted communicators. Before seeking to recreate such authority, however, it is important to understand how and why such authoritative sources of information emerged in the pre-internet era – which is to say, during the first six or seven decades of the twentieth century – when modern expectations about trust and a factual consensus developed.

Who were the creators and designators of trust during this period? In short, it was the institutional media. Moreover, through most of the twentieth century, institutional media acted as the gatekeepers of knowledge and news as well. Just who constituted the institutional media gatekeepers, however,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

changed over time. During the first part of the century, perhaps the crucial period in the development of gatekeepers and trusted communicators, it was major daily newspapers, especially those associated with William Randolph Hearst and Joseph Pulitzer, as well as Adolph Ochs's *New York Times*. As we shall discuss in more detail, in many ways it was cultural clashes between Hearst and Pulitzer on one side and Ochs on the other that generated the dominant gatekeeper/trusted-communicator model.[20]

After the First World War, while newspapers certainly maintained their importance, commercial radio broadcasters emerged as another crucial – and soon more popularly accessible – media institution. The first commercial radio station began broadcasting in 1920 in Pittsburgh, Pennsylvania. Four years later, 600 commercial radio stations were broadcasting in the United States. In 1926, the first national radio network, National Broadcasting Company (NBC), was formed.[21] As evidenced by President Franklin Delano Roosevelt's fireside chats during the Great Depression, radio quickly emerged as a widely available, popular means for institutional media – and those trusted communicators to whom they provided airtime, such as FDR – to reach mass public audiences.

Finally, around the mid century, at the beginning of what many considered the Golden Age of the institutional media, television broadcasters began to complement and eventually supplant radio (and newspapers) as the key institutional media. The Federal Communications Commission (FCC) first authorized commercial television broadcasts in 1941, but because of World War II, commercial television broadcasts did not begin in earnest until 1947.[22] And then the industry exploded. From 1946 to 1951, the number of television sets in use in the US rose from 6,000 to 12 million. By 1955, half of American households owned television sets.[23] Moreover, during the 1940s, the three iconic national television networks – NBC (evolved from the first radio network), Columbia Broadcasting System (CBS) (evolved from a competing radio network), and American Broadcasting Company (ABC) (spun off from NBC by order of the FCC) – had also emerged.[24] Finally, with the creation in 1956 of NBC's *The Huntley-Brinkley Report* (the first national television news broadcast), television's dominance as the primary source of news

[20] *See generally* W.Joseph Campbell, The Year that Defined American Journalism: 1897 and the Clash of Paradigms (2006).

[21] *KDKA Begins to Broadcast: 1920*, PBS (1998), https://www.pbs.org/wgbh/aso/databank/entries/dt20ra.html.

[22] Mitchell Stephens, *History of Television*, Grolier Encyclopedia, https://stephens.hosting.nyu.edu/History%20of%20Television%20page.html.

[23] *ibid*.

[24] *ibid*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

for most Americans (and the concomitant decline in the influence of newspapers) began.[25]

The rise of broadcasting also led to the emergence of the quintessential trusted communicators of this era, the network reporter and, later, anchorman. Coincidentally, the figures that epitomize both roles were affiliated with CBS. Edward R. Murrow first rose to prominence during the radio era through his revolutionary reporting on Hitler's *Anschluss* of Austria in 1938, and he became a household name by reporting live from London during the London Blitz in the early 1940s. He then moved to television and demonstrated continuing enormous influence through broadcasts, including a pathbreaking one in 1954 criticizing Senator Joseph McCarthy's witch hunt against Communists, which contributed to McCarthy's downfall.[26]

The other, even more important trusted communicator of the broadcast era was of course Walter Cronkite. Cronkite first became prominent (among other things, as the first designated "anchorman") during CBS's coverage of the 1952 presidential nominating conventions. But it was with the launch of *The CBS Evening News with Walter Cronkite* in 1962 that Cronkite's central role as *the* trusted communicator emerged.[27] Cronkite's influence was most famously demonstrated when his critical coverage of the Vietnam War in 1968 led to an important swing in public opinion against the war, and contributed to President Lyndon Johnson's decision not to run for reelection. Cronkite's status is illustrated by the fact that a 1972 poll named him "the most trusted man in America."[28] The institutional media and its key figures, epitomized by Murrow and Cronkite, were thus *the* trusted communicators of this era.

Even though their technology and reach varied, the gatekeepers/trusted communicators described earlier shared some basic characteristics. First, they were relatively scarce. The economics of newspapers meant that during most of this period, metropolitan areas could only support one or a handful of major newspapers.[29] With respect to the broadcast medium, the number of radio and television stations in any particular locality that actually produced original content (as opposed to playing music or broadcasting reruns of sitcoms) was limited by the same economic factors (essentially economies of scale) as newspapers. In addition, the fact that the number of possible

[25] *ibid.*

[26] David Mindich, *For Journalists Covering Trump, a Murrow Moment*, Colum. Journalism Rev. (July 15, 2016), www.cjr.org/analysis/trump_inspires_murrow_moment_for_journalism.php.

[27] Stephens, *supra* n. 22.

[28] *Walter Cronkite: American Journalist*, Britannica (Mar. 7, 2022), www.britannica.com/biography/Walter-Cronkite.

[29] *See* Miami Herald Pub'g Co. v. Tornillo, 418 U.S. 241, 249–50 and n.13 (1974).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

broadcast frequencies was physically limited – electromagnetic spectrum, as the Supreme Court has put it, is a "scarce resource"[30] – necessarily limited the number of broadcast outlets in any particular market. Indeed, in practice, the broadcast television market, especially in its role as disseminator of national news and general knowledge, was completely dominated by the three major networks (NBC, CBS, and ABC) until the launch of the Fox network in 1986 – and that only added one additional player. This situation only changed with the spread of cable television in the 1980s (and thus the end of spectrum scarcity because of the large channel capacity of cable systems), resulting in the launch of cable-only CNN in 1980 and then of Fox News in 1996.

The second shared characteristic between different types of gatekeepers and trusted communicators was that these gatekeepers sought to construct an "objective," nonpartisan image. The roots of this development, which has become an essential element of modern journalistic ethics,[31] can be found in the conflict between the sensationalist journalism championed by newspaper tycoons William Randolph Hearst and Joseph Pulitzer, and the "counteractivist," nonpartisan model of Adolph S. Och's *New York Times* (which Och purchased in 1896[32]). While the Hearst/Pulitzer model was dominant in the late nineteenth and early twentieth centuries, Ochs's commitment "to give the news impartially, without fear or favor, regardless of party, sect, or interests involved," a commitment Ochs announced on his first day of ownership of the *Times*,[33] eventually won out.[34] By 1920, this norm of objectivity[35] (which had previously gone by the name of "realism"[36]) was becoming the dominant paradigm of journalism, as reflected by the fact that the Society of Professional Journalists' first Code of Ethics, adopted in 1926, calls for journalistic "impartiality," meaning that "[n]ews reports should be free from opinion or bias of any kind."[37]

[30] Red Lion Broad. Co. v. FCC, 395 U.S. 367, 391 (1969).

[31] *See SPJ Code of Ethics*, Soc'y Prof. Journalists, www.spj.org/ethicscode.asp ("Ethical journalism should be accurate and fair").

[32] Bill Kovach and Tom Rosenstiel, The Elements of Journalism: What Newspeople Should Know and the Public Should Expect 76 (4th ed. 2021).

[33] *ibid*.

[34] *See generally* Campbell, *supra* n. 20; *Invisible Men: The Future of Journalism*, the Economist, July 18, 2020, at 67.

[35] Andrew Porwancher, *Objectivity's Prophet: Adolph S. Ochs and the New York Times*, 36 Journalism Hist. 186, 187 (2011), www.americanpressinstitute.org/journalism-essentials/bias-objectivity/lost-meaning-objectivity/.

[36] Walter Dean, *The Lost Meaning of "Objectivity,"* Am. Press Inst., https://perma.cc/6CRR-EWWL.

[37] *Sigma Delta Chi's New Code of Ethics*, Soc'y Prof. Journalists, http://spjnetwork.org/quill2/codedcontroversey/ethics-code-1926.pdf.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

It is important to note, however, that this goal of objectivity was a historical anomaly. Prior to the early twentieth century, newspapers and publishers did not pretend to be objective – to the contrary, they were explicitly partisan. Important historical examples include *The Aurora*, the newspaper edited by Benjamin Franklin Bache (Ben Franklin's grandson) in the late 1790s, which was tied to the Democratic-Republic Party of Jefferson and Madison (Bache and other Jeffersonian newspaper editors were prosecuted by the Adams Administration for sedition),[38] and Horace Greeley's *New York Tribune*, which was closely associated with the Republican Party before and during the Civil War.[39] Needless to say, these newspapers were *not* viewed as trustworthy by their political opponents (as demonstrated by Bache's prosecution). After World War I, however, economic pressures led to the consolidation of newspapers and a notable decrease in the number of daily newspapers – as epitomized by the merger in 1924 of the old rivals the *New York Herald* (which, though claiming to be nonpartisan, often supported Democratic Party policies during the Civil War) and Greeley's *New York Tribune*.[40] As a consequence, newspapers began to seek broader (and so bipartisan) audiences, which required them to abandon their partisan affiliations. Not coincidentally, journalistic ethics during this period also embraced objectivity as a desirable norm, as noted earlier.

The trend toward objectivity continued as newspapers were gradually supplanted by broadcast media: first radio, then (even more dominantly) television. For television broadcasting in particular, the push for objectivity was driven by similar economic motivations to maximize audience share because of the effective monopoly on national news held by the three national networks. In addition, the FCC's Fairness Doctrine, in effect from 1949 to 1987, strongly incentivized objectivity on the part of both radio and television broadcasters by requiring them to present opposing views on public issues, and by creating a right of reply on the part of individuals subject to a "personal attack" during broadcast programming.[41] Facially objective news coverage avoided triggering either requirement.[42]

[38] For a good discussion of this episode, see Geoffrey R. Stone, Perilous Times: Free Speech During Wartime 35 (2004).

[39] James M. McPherson, Battle Cry of Freedom: The Civil War Era 251–52 (1988).

[40] *New York Herald: American Newspaper*, Britannica, www.britannica.com/topic/New-York-Herald.

[41] Red Lion Broad. Co. v. FCC, 395 U.S. 367, 375–79 (1969); Matt Stefon, *Fairness Doctrine*, Britannica, www.britannica.com/topic/Fairness-Doctrine.

[42] The FCC itself, when it repealed the Fairness Doctrine in 1987, recognized that "the fairness doctrine provides broadcasters with a powerful incentive not to air controversial issue

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

This performed objectivity, playing out in a highly concentrated broadcast market, enabled a small set of individuals and institutions to emerge as "trusted communicators" in the eyes of a broad swath of the American public. We might call this the Murrow–Cronkite Effect. Furthermore, this institutional structure permitted trusted media figures to extend public trust to elite, designated "experts" *outside* the media by giving those experts the gatekeepers' imprimatur in the form of interviews and airtime (as an example, consider Edward R. Murrow's famous 1955 interview of Jonas Salk, the inventor of the polio vaccine[43]). As a consequence, during this "golden era," most of American society obtained news and knowledge from a few common and generally trusted sources.

What engendered this broad-based trust,[44] which in today's fractured world seems inconceivable? I would argue that the answer, in short, was a lack of alternative voices. The public trusted media gatekeepers because they had no choice – there were no significant opposing voices to question or undermine that trust because of concentration within the institutional media. It was precisely these factors – concentration and lack of choice – that made the institutional media, especially the three television networks, gatekeepers who exercised effective control over the flow of information into almost every American household. Indeed, it is hard to imagine how a media institution could play gatekeeper without this kind of option scarcity.

Furthermore, for economic reasons discussed earlier, these gatekeepers adopted an "objectivity" that overwhelmingly tended to reflect the views of the political center, in order to maximize their potential audience. As a consequence, there were simply no opportunities for the public to question consensus facts, or to become aware of what the institutional media was not telling them (such as President Kennedy's philandering, or the CIA's secret foreign coups during President Eisenhower's Administration). I am not insinuating that Murrow and Cronkite did not earn the public's trust – I have no doubt that they did, through ethical and insightful journalism. But that trust ultimately depended on a lack of access to alternative, non-mainstream voices.

Eventually, of course, this system of institutional concentration and consensus collapsed. The first developments along these lines are probably traceable

programming." *In re Complaint of Syracuse Peace Council against Television Station WTVH Syracuse, New York*, FCC 87-266, 2 FCC RCD. 5043, 5049-50 (1987).

[43] Michael Hiltzik, *On Jonas Salk's 100th Birthday, a Celebration of His Polio Vaccine*, L.A. TIMES (Oct. 28, 2014), www.latimes.com/business/hiltzik/la-fi-mh-polio-vaccine-20141028-column.html.

[44] To be fair, it is far from clear that the trust I am describing here extended to minority communities, but that is another story … Thanks to Helen Norton and Erin Carroll for (independently) pointing this out to me.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

to the FCC's repeal of the Fairness Doctrine in 1987,[45] which in turn led to the rise of right-wing talk radio, a medium which did not pretend or aspire to objectivity.[46] In addition, the explosion of the cable television medium during the 1980s ended the era of television concentration, because television no longer required scarce electromagnetic spectrum.[47] This in turn permitted the launch of the overtly partisan Fox News in 1996,[48] at the very dawn of the internet era. But while these developments began undermining the era of (supposed) media objectivity and the media's gatekeeper function, there can be little doubt that the internet, and especially the rise of social media, put a final end to the institutional media's control over public discourse. These, however, are relatively recent events. Twitter/X was founded in 2006,[49] the same year that Facebook became available to the general public.[50] But at first, these were relatively obscure platforms. It was not until the availability and widespread adoption of smartphones – the first iPhone was not released until 2007,[51] and smartphones did not come into common use for several years after then – that social media became mobile and easily usable, leading to its exponential growth.[52]

By the 2010s, the importance of social media in displacing traditional media as the primary engine of public discourse was evident – so much so that by 2017, that most hidebound of American institutions, the United States Supreme Court, recognized social media as "the most important places … for the exchange of views."[53] Every citizen became a potential publisher and people suddenly possessed a plethora of choices regarding what voices

[45] Stefon, *supra* n. 41.

[46] It is no coincidence that *The Rush Limbaugh Show* was launched nationally in 1988. *America's Anchorman*, Rush Limbaugh Show, www.rushlimbaugh.com/americas-anchorman/.

[47] During the 1980s, the number of cable networks exploded from 28 to 79, and cable penetration in American households enjoyed similar growth. *See* Brad Adgate, *The Rise and Fall of Cable Television*, Forbes (Nov. 2, 2020), www.forbes.com/sites/bradadgate/2020/11/02/the-rise-and-fall-of-cable-television/?sh=4b6145b76b31.

[48] Michael Ray, *Fox News Channel*, Britannica (Mar. 2, 2022), www.britannica.com/topic/Fox-News-Channel.

[49] Jack Meyer, *History of Twitter: Jack Dorsey and the Social Media Giant*, The Street (Jan. 2, 2020), www.thestreet.com/technology/history-of-twitter-facts-what-s-happening-in-2019-14995056.

[50] *Who We Are*, Meta, https://newsroom.fb.com/company-info/.

[51] *Apple Reinvents the Phone with iPhone*, Apple (Jan. 9, 2007), www.apple.com/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/.

[52] As an illustration, from 2008 to 2012, the number of Facebook users grew from 100 million to 1 billion – the latter being greater than the combined populations of the United States and the European Union. Kurt Wagner and Rani Molla, *Facebook's First 15 Years Were Defined by User Growth*, Vox (Feb. 5, 2019), www.vox.com/2019/2/4/18203992/facebook-15-year-anniversary-user-growth.

[53] Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (2017).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

to pay attention to, ending once and for all the gatekeeper function of the institutional media. And for the same reason, the range of opinions expressed publicly became massively more diverse, ending the media's role in creating consensus around a common set of facts and beliefs. The Murrow–Cronkite Effect had vanished.

With the collapse of the gatekeeper function also came the collapse of trusted communicators. There are no Edward Murrows or Walter Cronkites in the social media/Fox News era; instead we have Tucker Carlsons and Robert F. Kennedy, Jrs. (Mr. Kennedy, the son of Bobby Kennedy, is an active anti-vaccine propogandist who ran for President in the 2024 election cycle, before ultimately endorsing Donald Trump and then serving as Trump's Secretary for Health and Human Services[54]). This development is frankly unsurprising if one accepts, as I argued earlier, that much of the public's trust during the Murrow-Cronkite era was a product of the institutional media's gatekeeper function. No more gatekeepers, no more trust.

To be fair, the elimination of gatekeepers is not the only development that has contributed to the loss of trusted communicators. Most obviously, political polarization has also played an important role. As many people have drifted into more radicalized political positions, they inevitably cease to trust the traditional trusted communicators of the center (or, more honestly, the center-left) that made up the institutional media. Individuals whose views sit in the far-right or far-left have no reason to trust institutional speakers such as *The New York Times* or CNN. But here, too, the loss of gatekeepers plays an important causal role. During the peak of the gatekeeper era, most people had no access or exposure to radical voices unless they actively sought them out – and such voices were, as a result, quite rare. Today, social media and other internet forums provide easy access to a vast range of viewpoints and alternative "facts," permitting individuals to trust whomever they please – usually voices that reinforce and intensify their existing views. Of course, there have always been radical movements and conspiracy theories, but the rapid spread and sheer scope of the QAnon conspiracy theory, for example, would not have been possible in the pre-internet era; its ideas would never have gotten past the gatekeepers.

[54] Adam Nagourney, *A Kennedy's Crusade against Covid Vaccines Anguishes Family and Friends*, N.Y. Times (Feb. 26, 2022); Rebecca Davis O'Brien, Simon J. Levien, and Jonathan Swan, *Robert F. Kennedy Jr. Endorses Trump and Suspends His Independent Bid for President*, N.Y. Times (Aug. 23, 2024); Sheryl Gay Stolberg, *Senate Confirms Kennedy, a Prominent Vaccine Skeptic, as Health Secretary*, N.Y. Times (Feb. 13, 2025), https://www.nytimes.com/2025/02/13/us/rfk-jr-hhs-senate-confirmation.html.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

5.3 THE NEW GATEKEEPERS?

The loss of faith in institutional elites, including the institutional media, and the resulting collapse of consensus has had profound consequences. Most fundamentally, the loss of gatekeepers and trusted communicators has either threatened or eliminated the possibility of an ideology-free consensus on even basic facts. For individual media consumers, ideology seems to play a heavy role in shaping factual perceptions, regardless of objective reality. As an example, consider the fact that, in 2016, 72 percent of Republicans expressed doubts about Obama's birthplace, despite his Hawaiian birth certificate being in the public record.[55] And this loss of what one might call "consensus reality" has created an intellectual atmosphere of existential angst in some elements of American society. This is most evident within the mainstream media (perhaps unsurprisingly), but it is also an important part of the dialogue in politics (mainly on the left) and in academia (almost definitionally on the left).

To be clear, there is no question that a lack of factual consensus has had negative social consequences. It has made compromise – or even dialogue – across partisan lines far more difficult. And as the United States' experience with COVID-19 demonstrates, it can lead to deeply irrational policy choices (both on the left and right, to be clear). But the intellectual angst that I describe is often expressed in an existential manner, as fear for the very survival of our society (caused by such factors as the false belief among many Republicans, fostered by President Trump and elements of the conservative media, that the 2020 presidential election was stolen from Trump[56]).

As we saw in Chapter 2, the practical ways in which these elements of society have operationalized their angst has been to place enormous amounts of pressure on social media platforms such as Facebook, Twitter/X, and YouTube to actively block (among other things) online falsehoods in order to recreate a consensus reality. These critics want social media platforms to become the *new* gatekeepers, replicating the role of the twentieth-century institutional media in deciding what information and sources of information the public should be exposed to. Their logic appears to be that because a small number of social media platforms now host such a large portion of public discourse, the owners and controllers of those platforms should therefore ensure that the

[55] Josh Clinton and Carrie Roush, *Poll: Persistent Partisan Divide over "Birther" Question*, NBC News (Aug. 10, 2016), www.nbcnews.com/politics/2016-election/poll-persistent-partisan-divide-over-birther-question-n627446.

[56] *See, e.g.*, Zachary Ross, *The Five Biggest Threats Our Democracy Faces*, Brennan Ctr. for Just. (Dec. 15, 2020), www.brennancenter.org/our-work/analysis-opinion/five-biggest-threats-our-democracy-faces.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

flow of information to individuals is accurate and "clean," just as twentieth-century institutional media entities did when they controlled a similar bottleneck position.

And in fact, given their dominant market positions, the "big four" owners of the key social media platforms on which political discourse occurs – essentially Meta (which owns Facebook, Instagram, and Threads), Twitter/X, Alphabet (formerly Google, which owns YouTube), and ByteDance (which owns TikTok) – might well jointly possess the power to shape discourse akin to the three broadcast television networks of the twentieth century. But *should* social media firms be in the business of screening out false information and determining who is and is not a trusted communicator? Leaving aside the question of whether this is even possible (does anyone believe that Mark Zuckerberg can replace Walter Cronkite as "the most trusted man in America"?), I will argue that they should not.

There are several reasons why social media firms are ill-suited to be effective gatekeepers (or, as Mr. Zuckerberg would have it, "arbiters of truth"[57]). First and foremost, they have no economic incentives to do so. The traditional institutional media emphasized their objectivity and sought to develop reputations as trusted gatekeepers because it was in their economic interest. Objectivity and trust increased viewership and market share. The same is not true with social media. Social media platforms do not suffer from scarcity, and so can serve up an almost unlimited variety of content. Consequently, their algorithms emphasize relevance to users, not truth. That is what increases engagement, and so profits. Asking for-profit companies to take on roles that they have no economic incentive to adopt strikes me as both dubious policy and likely futile.

Second, social media firms have absolutely no expertise or training that would enable them to be either effective gatekeepers or effective identifiers of trusted communicators. As a practical matter, while social media algorithms are quite effective at sorting by relevance and interest, I am doubtful that they can be designed to identify "truth" or its opposite, given the tenuous and disputed nature of truth. More fundamentally, the people who work for the large tech firms are unlikely to be effective at the gatekeeper function. They are, after all, software engineers, not journalists or trained experts on subject matters such as science, history, or economics, and it seems unlikely, given the culture of Silicon Valley, that they will become so. Training the Mark

[57] Yael Halon, *Zuckerberg Knows Twitter for Fact-Checking Trump, Says Private Companies Shouldn't Be "The Arbiter of Truth,"* Fox News (May 27, 2020), www.foxnews.com/media/facebook-mark-zuckerberg-twitter-fact-checking-trump.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

Zuckerbergs of the world to be journalists is likely to be about as successful as it would have been to train Walter Cronkite to code.

Furthermore, social media platforms do not themselves generate content, unlike many traditional experts, which significantly reduces the incentives for these firms to develop serious in-house expertise (or for highly qualified experts to want to work for them – fact-checking is boring compared to content creation). Moreover, recent history suggests that when social media firms do rely on "expert" elites to identify misinformation, the results can be dicey – as illustrated by the fiascos of originally labeling the lab-leak theory of COVID-19's origins as misinformation,[58] or the decision to suppress a negative story about Hunter Biden and his laptop on the eve of the 2020 presidential election.[59] Indeed, social media critics are notably vague about how exactly social media firms are to identify "truth" (or its opposite, misinformation) going forward … other than, that is, strongly suggesting that misinformation is whatever they themselves – the political and media elites – deem it to be.

Finally, I would question whether *any* gatekeepers of information and/or "trusted communicators" are ultimately beneficial to society or consistent with principles of free expression. First, it is important to acknowledge that "truth," especially ideologically tinged truth, is a slippery thing.[60] While I do not deny the existence of objective facts (e.g., COVID-19 is real, and vaccines do work and do not cause autism), that sort of objectivity falls apart very soon after one gets beyond simple, provable facts. Certainly, COVID-19 is a real and dangerous disease, but where did it originate? Maybe a lab in Wuhan, maybe not – we may never know. Was closing primary schools for lengthy periods of time necessary to combat the spread of COVID-19? Teachers and parents may have different answers. Is it necessary or wise to vaccinate young children against COVID-19, given their low risk of severe illness? The experts-provided answers to these questions are, in truth, guesswork or opinions (albeit informed ones) dressed up as objective fact (or "science"). Should disagreement with these experts really be suppressed or labeled as misinformation? One would think not, even though that is precisely what progressive critics seem to be after.

The more fundamental question, once we get beyond a very narrow range of objective facts, is whether gatekeepers and deference to designated "experts" (i.e., trusted communicators) really offer the best way to identify "truth" and,

[58] *See* Brett Stephens, *Media Groupthink and the Lab-Leak Theory*, N.Y. Times (May 31, 2021).

[59] Andrew Prokop, *The Return of Hunter Biden's Laptop*, Vox (Mar. 25, 2022), www.vox.com/22992772/hunter-biden-laptop.

[60] For a thoughtful, extended discussion of this problem, see Jane Bambauer, *Snake Oil Speech*, 93 Wash. L. Rev. 73 (2018).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

conversely, misinformation. Those who favor gatekeepers, including social media gatekeepers, assume that gatekeepers and experts are necessary to hold back the tide of fake news. But there is a deep tension between this institutional approach and basic theories of free speech, as most famously encapsulated by Justice Oliver Wendell Holmes's foundational metaphor of the "marketplace of ideas": "that the best test of truth is the power of the thought to get itself accepted in the competition of the market."[61] Nor is it consistent with Justice Louis Brandeis's equally fundamental adage that, when faced with false or dangerous speech, "the remedy to be applied is more speech, not enforced silence."[62]

Both Holmes's and Brandeis's theories of free speech, while differing in details, are premised on the assumption that citizens should be permitted to freely engage in political debate, to the point even of advocating lawless behavior. This is because, according to Holmes, only then can truth emerge, and, according to Brandeis only then can citizens fully engage in our democracy. The concept of gatekeepers is simply inconsistent with both these visions. Gatekeepers are anathema to competition, and they are also quintessential silencers rather than enablers of "more speech."

Put differently, the gatekeeper solution advanced by progressive critics of social media, whereby a handful of elite actors control public discourse, is not consistent with either principles of free expression or the role of citizens in our democracy. Instead of trying to recreate a bygone (and, frankly, deeply flawed) era, perhaps we should be thinking about how to reinvigorate a marketplace of ideas and encourage genuine democratic deliberation that both surmount political polarization.

5.4 HARASSMENT AND HATE SPEECH

Aside from mis- and disinformation, the other type of speech regarding which progressives have been most critical of platforms is hate speech, as exemplified by the letter to Facebook from the group of Attorneys General discussed at the start of this chapter. Hate speech, which is to say attacks on groups based on protected characteristics, can take the form of general, political invective, or harassing speech directed at individuals. Either way, such speech, so long as it is online,[63] retains constitutional protections unless it rises to the level of an actual

[61] Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).
[62] Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring).
[63] If a racial or other epithet were hurled at an individual in person, it would almost certainly constitute unprotected "fighting words," but the fighting words category is limited to in-person insults and so does not apply to online speech. Cohen v. California, 403 U.S. 15, 20 (1971).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

threat of harm (it should be noted, however, that threats remain unprotected speech, even if the speaker has no intention of carrying out the threat[64]).

Despite the fact that hate speech enjoys constitutional protection, however, all the major social media platforms maintain and enforce bans on hate speech, either in the form of broad political statements or personal attacks.[65] It is noteworthy that Twitter/X retained its ban on hate speech after its purchase and renaming by Elon Musk, despite Musk's claims that he would prioritize free speech (though the actual incidence of hate speech on Twitter/X appears to have increased significantly after Musk's takeover, due to a dramatic reduction in company resources dedicated to content moderation[66]). Of course, progressive critics such as the state Attorneys General regularly criticize platforms for failing to adequately enforce their hate speech policies. For example, one prominent critic and activist described Facebook's response to complaints in this regard as "gaslighting."[67] But there is little doubt that the platforms seek to restrict hate speech (or in the case of Twitter/X under Musk, refuse to amplify it), regardless of how effectual their efforts are.

That all of the major platforms are willing to invest substantial resources – though perhaps Twitter/X less so than the others – to limit hate speech may seem surprising, but it shouldn't be. The reason is simple economics. The goal of social media platforms is to maximize user engagement, because engaged users can be sold as advertising targets. But realistically, the vast majority of users are repulsed by outright hate speech, and so the prevalence of such speech is likely to drive users away. And even more fundamentally, advertisers (who, after all, are the actual and ultimate customers of the platforms) most definitely do not want to be associated with hate speech, or for that matter any content that will drive away their own potential customers. In the jargon, advertisers are committed to "brand safety."[68]

[64] Virginia v. Black, 538 U.S. 343, 359–60 (2003).

[65] *See Hate Speech*, Meta, https://transparency.fb.com/policies/community-standards/hate-speech/ (current as of July 10, 2023) (Facebook); *Hate Speech Policy*, YouTube Help, https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436 (current as of July 10, 2023) (YouTube); *Hate Speech and Hateful Behaviors*, TikTok Community Guidelines (March 2023), www.tiktok.com/community-guidelines/en/safety-civility/#2 (TikTok); *Hateful Conduct*, Twitter, https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy (April 2023) (Twitter/X).

[66] Christian Martinez, *One Billionaire Owner, Twice the Hate: Twitter Hate Speech Surged with Musk, Study Says*, L.A. Times (April 27, 2023), www.latimes.com/business/technology/story/2023-04-27/hate-speech-twitter-surged-since-elon-musk-takeover.

[67] Charlie Warzel, *When a Critic Met Facebook: "What They're Doing Is Gaslighting,"* N.Y. Times (July 9, 2020), www.nytimes.com/2020/07/09/opinion/facebook-civil-rights-robinson.html.

[68] Vikram David Amar and Ashutosh Bhagwat, *Why Elon Musk's (and X's) Lawsuit against Companies Who Have Stopped Advertising on the X Platform Is Weak*, Verdict: Legal Analysis

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

The reality of this phenomenon is demonstrated by the fact that from April–May of 2022 to April–May of 2023, Twitter/X's advertising revenues declined by 59 *percent*, almost certainly as a result of Twitter/X's purchase by Elon Musk in the fall of 2022 and the subsequent rise in the prevalence of hate speech and pornography on the platform.[69] At least one CEO of a major advertiser who pulled ads from Twitter/X, Mondelez International (the maker of Oreos), publicly acknowledged that the reason was concerns about their ads appearing in proximity to "wrong messages" such as hate speech.[70] Indeed, the phenomenally fast uptake of Threads, the Meta/Facebook/Instagram app which was launched in July of 2023 to compete with Twitter/X (it shattered all previous records on download rates[71]), is almost certainly attributable to similar concerns among users.

Given the obvious incentives that platforms face, and react to, regarding hate speech (aside, perhaps, from the mysteries of Twitter/X under Elon Musk), why does so much hate speech remain available on platforms? Critics claim it is a lack of commitment, but this seems rather implausible given both economic incentives and the sheer scale of content moderation efforts by platforms. For example Facebook, traditionally the most criticized platform, employs both artificial intelligence and an army of 15,000 human content moderators globally to moderate content.[72] Rather, a simpler and more plausible explanation is the sheer scale of social media, combined with the difficulty of clearly defining what constitutes hate speech.

The scale is, of course, familiar. Facebook, Instagram, WhatsApp, YouTube, and TikTok all have over a billion monthly users (over two billion in the case of Facebook), producing an extraordinarily high number of daily posts. The simple,

and Commentary from Justia (Aug. 26, 2024), https://verdict.justia.com/2024/08/26/why-elon-musks-and-xs-lawsuit-against-companies-who-have-stopped-advertising-on-the-x-platform-is-legally-weak.

[69] Ryan Mac and Tiffany Hsu, *Twitter's U.S. Ad Sales Plunge 59% As Woes Continue*, N.Y. Times (June 5, 2023), www.nytimes.com/2023/06/05/technology/twitter-ad-sales-musk.html.

[70] Sheila Dang, *Analysis: Twitter's Advertising Business Seen Facing Slow Recovery*, Reuters (April 13, 2023), www.reuters.com/technology/twitters-advertising-business-seen-facing-slow-recovery-2023-04-13/; For an analysis of litigation arising out of advertisers' efforts to impose brand safety standards on a post-Musk Twitter/X, *see* Amar and Bhagwat, *supra* n. 68.

[71] Jay Peters and Jon Porter, *Instagram's Threads Surpasses 100 Million Users*, The Verge (July 10, 2023), www.theverge.com/2023/7/10/23787453/meta-instagram-threads-100-million-users-milestone. The more recent emergence of Bluesky as a potent Twitter/X alternative reflects similar trends. Callie Holtermann, *With Surge in New Users, Bluesky Emerges as X Alternative*, N.Y. Times (Nov. 12, 2024), www.nytimes.com/2024/11/12/style/bluesky-users-election.html.

[72] *How Does Facebook Use Artificial Intelligence to Moderate Content?* Facebook Help Center, www.facebook.com/help/1584908458516247; David Pilling and Madhumita Murgia, *"You Can't Unsee It": The Content Moderators Taking on Facebook*, Financial Times (May 17, 2023), www.ft.com/content/afeb56f2-9ba5-4103-890d-91291aea4caa.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

practical barriers to reviewing this volume of material is obvious. It is true that algorithms and artificial intelligence can generally identify the most obvious forms of hate speech such as well-known epithets (though it should be noted that even the most odious such epithets, such as the N-word, are not always used as hate speech). But on the margins, what constitutes "direct attacks against people – rather than concepts or institutions – on the basis of … protected characteristics" (Facebook's definition of hate speech) is not always easy to tell. Does a religiously based condemnation of homosexual conduct qualify as hate speech? Or what about a condemnation of such a religiously based condemnation?

The truth of the matter is that given the inevitable errors generated by scale, and the difficulty of defining hate speech at the margins, platforms will inevitably either under- or over-enforce their prohibitions on hate speech. In the United States, the perception on the political left is that platforms underenforce. But of course the perception on the political right is exactly the opposite; and in practice, no doubt both are true, in that sometimes hate speech slips through, and sometimes content that is not hate speech is blocked or mislabeled. Whether platforms (other than Twitter/X) are systematically under- or over-enforcing is therefore impossible to tell, and conclusions in that regard are more likely driven by a priori assumptions about what constitutes the "right amount" of content moderation than any real data.

Unlike in the United States, in Germany (and the European Union more generally), hate speech does not enjoy protected status. As a consequence, Germany adopted legislation commonly known as the NetzDG, effective January 1, 2018, imposing severe restrictions on online hate speech. Under that law, websites that do not, within twenty-four hours, remove hate speech that is "obviously illegal" under German law are subject to fines of up to fifty million euros.[73] The impact of this law is telling. While it successfully incentivized the platforms to restrict hate speech, commentators convincingly argue the law has also led to deletions of legitimate posts, and the chilling of political speech.[74] And worse, nations with less liberal agendas than Germany's have adopted copycat laws with the predictable result of significantly chilling or silencing legitimate speech the government disapproves of.[75] There are thus clear downsides to placing too much (whatever that means) legal or political pressure on platforms to increase content moderation of hate speech.

[73] *Germany Starts Enforcing Hate Speech Law*, *supra* n. 17.

[74] Rebecca Zipursky, Note, *Nuts About NETZ: The Network Enforcement Act and Freedom of Expression*, 42 Fordham Int'l L.J. 1325, 1359–60 (2019); Linda Kinstler, *Germany's Attempt to Fix Facebook Is Backfiring*, The Atlantic (May 18, 2018), www.theatlantic.com/international/archive/2018/05/germany-facebook-afd/560435/.

[75] Zipursky, *supra* n. 74, at 1360–62.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

Finally, what about speech that even in the United States lacks constitutional protection, such as true threats (or for that matter child pornography)? All the major platforms unsurprisingly ban such speech (for reasons discussed at the end of this chapter), but should we hold platforms legally responsible when they fail to block threats? Current law, in the form of Section 230 of the Communications Decency Act, immunizes platforms, but should we change that? The answer, I would argue, is no. The reality is that even assuming that platforms are making good faith efforts to block personal threats, their efforts will inevitably be imperfect.

Consider the phrase "I'm going to kill you." Those words are uttered, written, and posted online thousands of times every day, and in almost every case the phrase at most is intended to convey anger, but usually is more of a joke. Yet precisely those words can, of course, constitute a serious threat of violence. Context is all, and most of us can, from context, easily figure out how the words are meant. But algorithms are terrible at context, and so will guess wrong with some frequency about the true meaning of such phrases. And because context is often cultural, even English-speaking human content moderators, if they were not raised in the United States, could easily misunderstand the intended meaning.

Now also consider the fact that the United States Supreme Court recently (in 2023) held that for speech to constitute an unprotected threat, the speaker must have acted recklessly – that is, "that a speaker is aware 'that others could regard his statements' as threatening violence and 'delivers them anyway.'"[76] How, exactly, is an algorithm (or even a distant human moderator) to determine the mental state of an individual who posts a potentially threatening statement online? Even after full, criminal trials mental states are notoriously difficult to pin down, so to expect content moderation, which happens constantly and necessarily quickly, to make such determinations is absurd.

Given the inevitability of mistakes, it becomes clear that just as NetzDG has led platforms to over-moderate potential hate speech in Germany, legal liability will cause platforms to block any speech that could plausibly be considered threatening. But as the "I'm going to kill you" example demonstrates, that is a huge amount of speech. Furthermore, given the vehemence of modern, online political discourse, a substantial fraction of that speech will be sharp political criticisms rather than actual threats. The line is often very difficult to draw, and if they face liability platforms will not take the risk. The burden on speech this would impose is, I would argue, simply not worth the marginal benefits of amending Section 230.

[76] Counterman v. Colorado, 600 U.S. 66, 79 (2023) (quoting Elonis v. U.S., 575 U.S. 723, 746 (2015) (Alito, J., concurring in part and dissenting in part)).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

5.5 POLITICAL MANIPULATION AND POLARIZATION

Political manipulation, such as Russian interference in the 2016 and 2020 presidential elections, and its sister phenomenon, political polarization, are unquestionably serious concerns. And it is possible that the rise of social media has enabled and increased both phenomena – though, as discussed in Chapter 2, the degree to which social media contributes to polarization is quite unclear. But to what extent can platforms be cajoled or forced to moderate manipulative or polarizing content?

As to political manipulation, all of the major platforms (except perhaps Twitter/X in the Elon Musk era) appear to be firmly committed to combating such efforts, and what evidence we have does suggest that the platforms were far more effective in combating manipulative content in 2020 and 2024 than in 2016. Nonetheless, the media and public opinion should certainly continue to press platforms to fight political manipulation. But as with so many things, difficulties in distinguishing between legitimate and illegitimate online strategies makes legal intervention essentially impossible, especially when manipulation consists of spreading true but divisive information.

None of which is to say that no action is possible. Platforms do, and undoubtedly will, continue to block bots and other fake accounts, especially those seeking to manipulate elections in the United States but originating from outside the country. Nor does official pressure to block such accounts raise serious constitutional concerns, because foreign actors outside the territory of the United States do not enjoy First Amendment rights.[77] But with respect to content that is domestically produced or distributed by legitimate users, there is in truth little to be done because such content constitutes legitimate, and constitutionally protected, political speech.

Polarization is, however, a very different matter. Polarization is not, primarily, a product of disinformation, or even of manipulation. It is rather an increasingly prevalent element of political culture in the United States (and elsewhere, though seemingly not to the same degree in most other countries). Social media platforms' role in stoking polarization is unclear but probably limited. Even if, as Jonathan Haidt argues,[78] social media algorithms do increase polarization, however, the ultimate cause and source of political

[77] Agency for International Development v. Alliance for Open Society International, Inc., 591 U.S. 430 (2020).

[78] Jonathan Haidt, *Why the Past 10 Years of American Life Have Been Uniquely Stupid: It's Not Just a Phase*, The Atlantic (April 11, 2022), www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/. For a collection of Haidt's arguments on this topic, *see* https://jonathanhaidt.com/social-media/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

polarization is not Facebook or Twitter/X. It is rather we, the users. And this fact poses a fundamental problem for those who would regulate social media to combat polarization.

Consider first the possibility that the pathway by which social media enhances political polarization is not via platform algorithms, but rather through users' choice of friends who share their political preferences, as (recall from Chapter 2) Professor Jane Bambauer and her co-authors argue.[79] If that is the case, then to reduce the polarizing impact of social media, either the state or platforms would have to force users to friend or follow individuals they disagree with, and as a result probably do not like. But the very thought of doing so seems incredibly manipulative, and ultimately frankly ridiculous. Coercing users in this way would deeply reduce people's enjoyment of social media and probably drive many users way. Perhaps some of the sharpest critics of social media would think this is a good outcome; but most of society should surely recoil at the thoroughly Luddite "solution" of destroying an important, new technological tool which has many, many legitimate uses (especially in commerce), and which most people enjoy – why, after all, would they spend so much time on social media if they did not.

But perhaps Professor Haidt is correct and the real source of increasing polarization is platform algorithms, which seek to maximize user engagement by serving up extremist or overwrought content. If that were the case, then legally requiring platforms to tweak their algorithms might reduce the tendency toward polarization (I say legally require because platforms, which are after all for-profit enterprises, are unlikely to voluntarily take steps that reduce engagement, and so advertising profits). The difficulty is that polarizing material is almost always fully legal and constitutionally protected. Indeed, because such content is typically political in nature, it sits at the very heart of public discourse and so constitutional protections. As such, legally restricting the spread or amplification of such content would violate the First Amendment rights of both users, and of platforms themselves. After all, a platform's decision to amplify specific content is itself an expressive act, as well as an editorial one as discussed in Chapter 4, both aspects of which receive First Amendment protections.[80]

But if legal requirements are a nonstarter, should the public and the media not seek to cajole, and the government jawbone, platforms into altering their algorithms? Perhaps, but even here it is important to take a pause. When

[79] Jane R. Bambauer, Saura Masconale, and Simone M. Sepe, *Cheap Friendship*, 54 U.C. Davis L. Rev. 2341 (2021).

[80] Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard*, 1 J. Free Speech L. 227 (2021).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

commentators criticize platforms for seeking to maximize engagement, consider what they are really saying. They seem to be saying that platforms should serve up content that users don't like, or like less, so that they spend less time on the platform. This is the equivalent of trying to convince restaurants to serve food that customers will not enjoy, so as to convince them to spend less time and money at restaurants. But even if the goal was less problematic – say, to have customers eat less fatty food – we still do not typically interfere with private choices in this way, or force businesses to take steps that will make customers unhappy. In other words, criticizing platforms for "maximizing engagement" is essentially criticizing them for providing content that users desire and prefer.

Finally, for all the noise about "addiction" and the like, social media platforms are hardly heroin. In our world of infinite entertainment and communications options from Netflix to texting, no one is forced to use social media even for entertainment. And when social media is used for other functions, such as commercial sales, no one thinks that problematic or socially destructive. In short, social media platforms, like most commercial businesses, provide the products that in their view will maximize sales. And absent evidence that the provided product causes severe harm, as in the case of opioids, in a capitalist society that is generally considered normal and desirable even if, as with fatty foods, alcohol, and (perhaps) social media, the desired product is not necessarily good for the consumer.

5.6 INCENTIVES, MOTIVES, AND THE CASE FOR HUMILITY

Where does all of this leave us? The answer is clearly not no content moderation, because the platforms themselves agree (for reasons touched on in Chapter 4) that some level of moderation is essential, both for societal good and their own business models. This fact, combined with the very different incentives and motives of private internet firms versus the government, leads me to conclude that platforms are best left to their own devices in creating and enforcing content moderation rules, within broad limits.

Let us start with government incentives. The starting point is the perhaps grim but inevitable fact that political leaders of all stripes like to stay in power. In democracies, that means winning elections. In autocracies, it means suppressing dissidents. But the goal remains the same; and this fact alone creates strong motivations for political leaders vis-à-vis free speech and platforms.

Let us first consider the motivations of democratically elected leaders. In democracies, free speech is foundational and essential. Without free speech, citizens cannot meaningfully discuss public policy or the achievements and failings of elected leaders, and so cannot cast their vote intelligently. And

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

more broadly, freedom of expression and related liberties such as assembly and association are the crucial, necessary tools with which citizens engage and communicate with their leaders. But from the point of view of elected officials, free speech is of course a threat, because it can be used to reveal their errors and weaknesses. Over time, this in turn can undermine support for them, and so their ability to prevail in elections. Hence the motivation to censor unfavorable speech. Of course, elected leaders must be careful in how they censor, or the censorship itself becomes a political problem, but so long as leaders target minority or unpopular viewpoints, they can often get away with suppression. After all, democratic leaders do not need universal support, just that of a majority of citizens. That is why constitutional protections for freedom of speech, ideally enforced against elected leaders by an independent judiciary, are an essential element of a successful democracy.

Now consider autocracies. Here, the motivation to suppress speech is even more obvious – speech is the primary and essential means to organize opposition to autocratic leaders. It is no coincidence that the largest and most successful autocracy in the world – China – also has the most elaborate and successful censorship systems. And unlike democratic leaders, autocratic leaders do not face democratic checks on their desire to censor.

Finally, consider the motivations of private social media platforms. Unlike government officials, at heart the goal of such firms is to maximize speech, because that is in some sense the product there are providing. To be more precise, platforms host speech to attract users, and then make money by selling access to those users to advertisers. Platforms cannot adopt aggressive rules restricting content because their financial goal is to maximize users; and to maximize users they need to host a great variety of speech that attracts a broad range of users, with a broad range of tastes. Furthermore, from the point of view of the platform, it is entirely irrelevant if the speech they host is favorable to the government, unfavorable to the government, or has nothing to do with government – the more the merrier.[81] Furthermore, even content which is

[81] The one caveat here is that if the government and/or political parties affiliated with the government are themselves a major source of platform profits, say from purchasing political advertising, then there might be occasions when platforms find it profitable to block anti-government speech in order to retain government business, to the detriment of other users. Such situations seem likely to be relatively uncommon, however, because political advertising constitutes a tiny fraction of overall advertising revenues for platforms – Facebook, for example, is the single largest conduit for online political ads, yet political advertising constituted less than 1 percent of company revenues. Katie Canales, *Mark Zuckerberg Said Facebook Makes a "Relatively Small" Amount from Political Advertising. The Company Has Made $2.2 Billion From Political Ads since Mid-2018)*, Business Insider (Oct. 28, 2020), www.businessinsider.com/zuckerberg-facebook-political-ad-revenue-2020-10.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006

unpopular with the majority of users typically is of interest to some elements of the population, and so to maximize users, platforms are incented to permit that speech. It is only when speech is so unpopular with users that it is likely to chase them off the platforms that platforms have will want to suppress it. That, then, is the role of content moderation policies: not to suppress speech broadly or to tilt the political dialogue. Rather, it is to suppress the worst of the worst – like terrorist propaganda, hate speech, threats, and (for some platforms such as Facebook) pornography – that is likely to repel significant numbers of users, while otherwise maximizing speech in order to maximize engagement and profits. This is unlike with governments, which even in democracies have no incentive to permit unpopular speech, since their interests are in pleasing the majority, not niche minorities.

These points may seem obvious, but they have an important implication. Contrary to current orthodoxy, we should be far more trusting of platforms restricting speech than politicians restricting either speech or platforms because platforms have no systematic anti-speech bias, but government officials most certainly do (at least as to speech critical of them). As a result, it is as predictable as the sun rising in the east that *anytime* a government regulates in the expressive sphere, including regulating platforms, one of the core purposes of the regulation will be to maximize speech favorable to the government, and to minimize speech unfavorable to it. This is obviously true in autocratic states like China; but it is also true of democratically enacted legislation such as the laws recently enacted in the US states of Florida and Texas, both of whose governors publicly admitted (indeed, emphasized) that the purpose of the laws was to enhance conservative voices (both governors are leading conservatives, and members of the Republican Party).[82]

In short, government remains a much greater threat to free speech than social media platforms, not only because of the former's monopolies on violence and control but also because of their perverse incentives. The primary motivation of internet companies, on the other hand, is to make money, which in the free speech sphere is actually quite innocuous – after all, that is the motivation that drives all privately owned media. So, just as we leave it to the owners of legacy media to decide what (legal) content they will publish, so too the best solution at hand may be to leave that power in the hands of the platforms.

[82] NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1205 (11th Cir. 2022); NetChoice, LLC v. Paxton, 1:21-CV-840-RP, 2021 WL 5755120, at *1 (W.D. Tex. Dec. 1, 2021).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 10:46:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.006