2

The Progressive War

Social Media as Enablers

If conservative criticisms of social media ultimately come down to the claim that social media platforms, through their content moderation practices, suppress too much speech, progressive claims come down to the assertion that platforms do not suppress enough harmful speech. Both sets of criticisms are, to some extent, sincere. And just as there is some basis for conservative unhappiness with platforms, so too there is some legitimacy to progressive claims. But the fact that both sets of polar-opposite criticisms can co-exist tells us something both about the extraordinary difficulty of achieving Goldilocks-style "just right" content moderation and about our broader social tendency to attribute fundamental dysfunctions rooted in our culture to social media.

2.1 POLITICAL MANIPULATION

Let us begin by laying out the basic elements of the progressive case. One specific area where platforms have been especially criticized is in their seeming inability to block misinformation, disinformation, and other forms of online political manipulation. These concerns first became prominent during the 2016 presidential election. After Donald Trump's victory in that close-fought election, evidence emerged that a number of actors, both foreign and domestic, engaged in a variety of forms of political manipulation on social media platforms, many (though not all) with the goal of benefiting the Trump campaign.[1] The tools used to manipulate voters in 2016 ranged from outright

[1] Robert Yablon, *Political Advertising, Digital Platforms, and the Democratic Deficiencies of Self-Regulation*, 104 MINN. L. REV. HEADNOTES 13, 14 and n.5 (2020) (citing Nathaniel Persily, *Can Democracy Survive the Internet?*, 28 J. DEMOCRACY 63, 67–71 (2017); Abby K. Wood and Ann M. Ravel, *Fool Me Once: Regulating "Fake News" and Other Online Advertising*, 91 S. CAL. L. REV. 1223, 1229–34 (2018)).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

disinformation – "fake news" – to more subtle attempts to increase social divisions on hot button political issues by using bots and other devices to spread stories quickly.[2] And while the actual impact of all of this on the election results is unknowable (though probably fairly small), it is no surprise that commentators and politicians on the political left were upset by these revelations.

Most prominent and controversial, and to this day disputed by Donald Trump and his supporters,[3] was evidence that the government of Russia engaged in a massive disinformation and manipulation campaign on social media during the 2016 election cycle, with the goal of benefiting Trump's campaign at the expense of both his Republican rivals during the primary season and his ultimate Democratic opponent Hilary Clinton.[4] One prominent example of such manipulation concerns the Internet Research Agency, a (now closed) Russian company owned by Yevgeny Prigozhin, then an ally of Russian President Vladimir Putin (Prigozhin later led the Wagner Group uprising against Putin's government, and then died in a "mysterious" airplane crash). During the 2016 election, the Internet Research Agency reportedly spread stories using fake accounts, claiming to be American, that supported candidate Trump's attacks on various individuals and government institutions. In addition, and most strikingly, evidence came to light that Russian manipulation was especially targeted at particular segments of African American voters, a demographic that tilts heavily Democratic, seeking to discourage the targets from voting for Hilary Clinton.[5] These tactics, which exploited real and existing racial tensions and grievances such as concerns about President Bill Clinton's record on race issues during his presidency, sought to convince voters to refrain from voting during the general election, or to support the Green Party candidate Jill Stein. Revelations such as these inevitably outraged political progressives. One concern of these critics was of course the political impact of the Russian actions. But more fundamentally, these actions, in seeking to suppress the African American vote, were deeply racist and as such come with a long and troubling historical pedigree.

In the wake of revelations about especially the Russian election interference, the major platforms took substantial steps to try and limit future

[2] Wood and Ravel, *supra* n. 1, at 1229–32.

[3] Julie Hirschfeld Davis, *Trump, at Putin's Side, Questions U.S. Intelligence on 2016 Election*, N.Y. Times (July 16, 2018), www.nytimes.com/2018/07/16/world/europe/trump-putin-election-intelligence.html.

[4] *Ibid.*; *see also* Derek E. Bambauer, *Information Hacking*, 2020 Utah L. Rev. 987, 987–94 (summarizing various Russia-backed disinformation campaigns in 2016).

[5] Scott Shane and Sheera Frenkel, *Russian 2016 Influence Operation Targeted African-Americans on Social Media*, N.Y. Times (Dec. 17, 2018), www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

manipulation.[6] Nonetheless, the US intelligence community documented strong evidence that Russia continued, in the 2020 election, to seek to benefit President Trump's reelection campaign in a number of ways, including via the spread of misinformation.[7] Such reports again, inevitably, triggered sharp complaints from Democratic political leaders, albeit perhaps not as sharp as in the wake of the 2016 election because of Democratic candidate Joe Biden's ultimate victory in the presidential election.

2.2 MISINFORMATION AND THE COVID-19 PANDEMIC

If election manipulation was the original trigger for progressive concerns about online disinformation and manipulation, the COVID-19 pandemic led to their apex.

As everyone knows, COVID-19, which was first detected in the Chinese city of Wuhan in December of 2019, was declared a global pandemic by the World Health Organization in March of 2020 (specifically, March 11). The ensuing stay-at-home orders, business closures, and other unprecedented measures created massive social and economic disruptions, the long-term effects of which are likely to be felt for decades. Conspiracy theories, a variety of unsupported factual claims, and some level of outright lies about the coronavirus causing the pandemic began to spread on social media almost immediately after the Wuhan outbreak, well before the WHO's March declaration. These included claims that the virus was deliberately engineered and released, and (my favorite for sheer wackiness) that 5G cellular networks helped create or spread the virus.[8]

Along with misinformation about the source and spread of COVID-19, early in the pandemic controversial stories spread about potential treatments for the virus. Most dangerously, a nontrivial number of individuals, some inspired by off-hand comments by President Trump in April of 2020, ingested bleach or other disinfectants as purported cures for COVID; the results were predictably tragic.[9] Less dangerous, though as noted in Chapter 1 also probably wrong,

[6] Ashutosh Bhagwat, *The Law of Facebook*, 54 U.C. Davis L. Rev. 2353, 2363–65 (2021).

[7] Julian E. Barnes, *Russian Interference in 2020 Included Influencing Trump Associates, Report Says*, N.Y. Times (March 16, 2021), www.nytimes.com/2021/03/16/us/politics/election-interference-russia-2020-assessment.html.

[8] Josh Taylor, *Bat Soup, Dodgy Cures and "Diseasology": The Spread of Coronavirus Disinformation*, The Guardian (Jan. 30, 2020), www.theguardian.com/world/2020/jan/31/bat-soup-dodgy-cures-and-diseasology-the-spread-of-coronavirus-bunkum.

[9] Nicholas Reimann, *Some Americans Are Tragically Still Drinking Bleach as a Coronavirus "Cure,"* Forbes (Aug. 24, 2020), www.forbes.com/sites/nicholasreimann/2020/08/24/some-americans-are-tragically-still-drinking-bleach-as-a-coronavirus-cure/?sh=110fb41b6748.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

were claims by President Trump beginning in May of 2020 that hydroxychloroquine, a malaria drug, was effective in treating COVID-19.[10] Then, in the summer of 2021 a number of prominent conservative media personalities (mainly associated with Fox News) began endorsing a story that had developed in certain social media circles (notably on the far-right platform Gab) that the drug ivermectin, a deworming medication primarily used for horses, could cure COVID-19[11] (the Food and Drug Administration (FDA), the US government agency responsible for regulating drug safety, prior to the second Trump Administration firmly opposed the use of ivermectin,[12] and later research suggests that ivermectin is not effective in treating the disease[13]).

For our purposes, what is important and interesting about COVID-19 mis- and disinformation was not so much the existence of the phenomenon, but rather the response of the major platforms to it, and the political dynamic that emerged from all of this. As for the platforms, very early in the pandemic, misinformation was rife across all social media platforms. Indeed, given the lack of information about or scientific understanding of the virus, it was often impossible in the early days to distinguish misinformation from guesswork, whether factually based or not. As an example of such shoddy, but no doubt sincerely intended, guesswork, consider the prediction in March of 2020 by well-known and respected law professor Richard Epstein (with whom I studied, full disclosure) that the total number of COVID deaths in the United States would be approximately 5,000 (increased from his original estimate of 500).[14] Given that as of April of 2023 the Centers for Disease Control and the World Health Organization both report over 1 million COVID deaths in the United States (a possibility that Epstein pooh-poohed in his essay), Epstein was obviously dead wrong. Furthermore, if his and other sanguine estimates led people to ignore safety warnings, they might have cost lives. But, given the

[10] Andrew Solender, *All the Times Trump Has Promoted Hydroxychloroquine*, Forbes (May 22, 2020), www.forbes.com/sites/andrewsolender/2020/05/22/all-the-times-trump-promoted-hydroxychloroquine/?sh=fd19820464 32; Katie Thomas, *F.D.A. Revokes Emergency Approval of Malaria Drugs Promoted by Trump*, N.Y. Times (June 15, 2020), www.nytimes.com/2020/06/15/health/fda-hydroxychloroquine-malaria.html.

[11] Oliver Darcy, *Right-Wing Media Pushed a Deworming Drug to Treat COVID-19 that the FDA Says Is Unsafe for Humans*, CNN (Aug. 23, 2021), www.cnn.com/2021/08/23/media/right-wing-media-ivermectin/index.html.

[12] The link to the relevant page on the FDA website appears to have been disabled by officials in the second Trump Administration (as of April 2025).

[13] Susanna Naggie et al., *Effect of Ivermectin vs Pacebo on Time to Sustained Recovery in Outpatients with Mild to Moderate COVID-10: A Randomized Clinical Trial*, JAMA Network (Oct. 21, 2022), https://jamanetwork.com/journals/jama/fullarticle/2797483.

[14] Richard A. Epstein, *Coronavirus Perspective*, Hoover Institute (March 16, 2020), www.hoover.org/research/coronavirus-pandemic.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

lack of information about COVID-19 available in March of 2020, such claims cannot be described as mis- or disinformation.

Claims regarding the origins of the virus, as well as claims to have discovered "miracle cures," were also, given the lack of data, extremely hard to evaluate or refute early in the pandemic (though even in the early days, it was surely clear that drinking bleach was a very bad idea, and that 5G networks did not create the virus). As such, the failure of platforms (and news organizations) to filter trustworthy from non-trustworthy information is understandable. Nonetheless, because of the risks associated with COVID misinformation, the major platforms began collaborating with outside experts to identify, label, and sometimes remove content determined to be misinformation. Facebook's and Twitter/X's first efforts in this direction, focused in particular on misinformation likely to lead to material harm, were adopted in January of 2020, but expanded significantly in later months, as more scientific consensus emerged regarding COVID.[15] YouTube adopted a similar policy in May of 2020,[16] and as of April of 2023 Meta (the owner of Facebook and Instagram) and Google (the owner of YouTube) continued to enforce their policies. Twitter/X, however, stopped enforcing its policy in November of 2022, in the wake of Elon Musk's purchase of the platform.[17]

The adoption of anti-misinformation policies did not, of course, eliminate the spread of falsehoods and conspiracies about COVID. For one thing, enforcement of misinformation policies was inevitably imperfect. A big part of the reason for this is that the sheer scale of social media makes perfect content moderation impossible, especially because the major platforms operate in many different languages. But in addition, the state of knowledge and the scientific consensus regarding COVID changed, making the accuracy or inaccuracy of some arguable claims (such as the efficacy of hydroxychloroquine as a COVID treatment) only clear over time. Indeed, in some cases it turned out that uncertainty led the major platforms to over-suppress, such as with the theory that the COVID-19 virus leaked from a research lab in Wuhan, China, a theory which was originally labeled misinformation but which later (in Facebook's case in May of 2021[18]) was permitted, in light of the ongoing uncertainty about the origins of the virus.

[15] Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation about COVID-19*, Meta Newsroom (April 16, 2020), https://about.fb.com/news/2020/04/covid-19-misinfo-update/. The relevant Twitter/X link has been disabled as of April 2025.

[16] *YouTube Help: Medical Misinformation Policy*, https://support.google.com/youtube/answer/9891785?hl=en.

[17] Associated Press, *Twitter Will No Longer Enforce Its COVID Misinformation Policy*, NPR (Nov. 29, 2022), www.npr.org/2022/11/29/1139822833/twitter-covid-misinformation-policy-not-enforced.

[18] Rosen, *supra* n. 15.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

In addition to the imperfections of content moderation on the major platforms, however, another important avenue for the spread of misinformation was the fact that smaller platforms such as Gab and Parler failed to adopt content moderation policies similar to those on the major platforms. As a consequence, these platforms became important avenues of misinformation, especially about vaccines, for individuals (admittedly a self-selecting minority) who frequented such sites.[19] Combined with the continuing spread of vaccine skepticism on more traditional media such as Fox News,[20] it seems clear that COVID misinformation contributed, to some significant but unknowable extent, to reduced rates of vaccination against COVID within the United States and elsewhere, thereby increasing sickness and mortality.

These events inevitably drew a sharp, critical response directed at the major social media platforms (and to a lesser extent, minor ones). What is striking, however, is that the attacks on social media platforms for permitting the spread of COVID misinformation (or more accurately, for not doing enough to prevent such spread) came almost exclusively from the political left. This is because public discussion of COVID origins and cures had, from the beginning, a sharp political divide. On the progressive left, however, the message was clear. In the summer of 2021 the Biden Administration, led by Surgeon General Vivek Murthy, strongly urged tech companies to fight vaccine misinformation, and President Biden went as far, in a press conference, as to accuse platforms of "killing people."[21] These and similar actions ultimately led to a lawsuit against the Biden Administration claiming that their pressure campaign violated the First Amendment. This lawsuit, though it met with initial success in the ultra-conservative United States Court of Appeals for the Fifth Circuit,[22] was ultimately dismissed by the US Supreme Court.[23]

Nor were congressional leaders from the Democratic Party silent on the matter. Tech-savvy members such as Senators Amy Klobuchar and Mark Warner cheered the Biden Administration's efforts. Separately, Senator

[19] Sheryl Gay Stolberg, *A Lasting Legacy of Covid: Far-Right Platforms Spreading Health Myths*, N.Y. Times (Nov. 22, 2022), www.nytimes.com/2022/11/22/us/politics/covid-misinformation-gab.html.

[20] Tiffany Hsu, *Despite Outbreaks among Unvaccinated, Fox News Hosts Smear Shots*, N.Y. Times (July 11, 2021), www.nytimes.com/2021/07/11/business/media/vaccines-fox-news-hosts.html.

[21] Rebecca Klar, *Feds Step Up Pressure on Social Media Over False COVID-19 Claims*, The Hill (July 18, 2021), https://thehill.com/policy/technology/563470-administration-puts-new-pressure-on-social-media-to-curb-covid-19/.

[22] Missouri v. Biden, 83 F.4th 641 (5th Cir. 2023).

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

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

Klobuchar (who represents Minnesota) and Senator Ben Ray Lujan of New Mexico introduced legislation that would have imposed liability on platforms if they became vehicles for the spread of medical misinformation (specifically, the legislation would have stripped tech companies of the immunity they normally enjoy under Section 230 of the Communications Decency Act for third party content posted on their platforms – Section 230 is discussed in more detail in Chapter 6).[24] Senator Elizabeth Warren of Massachusetts similarly issued strong and sharply critical comments about the role of tech platforms in spreading misinformation, albeit she did not focus on medical misinformation specifically.[25] And more generally, Democratic politicians throughout the country urged social media platforms to curb COVID misinformation – indeed, the Democratic leadership in California went so far as to pass legislation that would have caused disciplinary sanctions to be imposed on doctors who spread COVID disinformation (defined as "false information that is contradicted by contemporary scientific consensus contrary to the standard of care").[26]

Finally, the message of traditionally left-leaning news outlets such as the *New York Times* and *Washington Post* was largely the same as that of progressive politicians. News stories largely supported the narrative of Big Tech incompetence and conservative involvement in the spread of misinformation, and editorials regularly flayed the major platforms for their inaction. On the mainstream political and journalistic left, it would seem that a wide consensus exists regarding the failure of social media platforms to control the spread of medical misinformation, and their moral (and perhaps legal) obligation to cure those failures.

In short, on the subject of social media platforms' efforts to control medical misinformation during the COVID-19 pandemic, progressives and conservatives have taken polar-opposite positions. Progressives have consistently, and sharply, criticized such efforts as inadequate, and expressed concerns about lost lives. Conservatives, on the other hand (as discussed in Chapter 1), have criticized them as excessive and have consistently expressed concerns about threats to liberty, going so far as to sue the Biden Administration over their

[24] *News Releases: Klobuchar, Luján Introduce Legislation to Hold Digital Platforms Accountable for Vaccine and Other Health-Related Misinformation* (July 22, 2021), www.klobuchar.senate.gov/public/index.cfm/2021/7/klobuchar-luj-n-introduce-legislation-to-hold-digital-platforms-accountable-for-vaccine-and-other-health-related-misinformation.

[25] *Fighting Digital Disinformation*, https://elizabethwarren.com/plans/fighting-digital-disinformation.

[26] Brendan Pierson, *California Law Aiming to Curb COVID Misinformation Blocked By Judge*, Reuters (Jan. 26, 2023), www.reuters.com/business/healthcare-pharmaceuticals/california-law-aiming-curb-covid-misinformation-blocked-by-judge-2023-01-26/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

efforts to pressure the platforms on these issues. As we shall see next, COVID is far from the only subject on which such a stark dichotomy has emerged.

2.3 MOB HARASSMENT, DOXING, AND THREATS

One of the most persistent, concerning, and impactful forms of online misbehavior is mob attacks on individuals, often via social media platforms. The forms of such mass harassment vary. One common form of harassment is doxing, whereby private information about individuals is released online, including such things as phone numbers, home addresses, and intimate images (often called "revenge porn," or – more accurately – nonconsensual pornography[27]). Threats of violence are also quite common, and especially likely to be directed at women. One particularly well-publicized example of this was the thoroughly awful "Gamergate" events in 2013–14, during which a succession of women involved in, or writing about, the video game industry were targeted by (overwhelmingly male) online mobs with horrifying threats of violence (and in particular, of rape, illustrating the heavily misogynistic nature of such attacks). These threats were sometimes accompanied by doxing of personal information such as home addresses. In one instance a women named Brianna Wu, who had cofounded an indie game studio, was targeted simply for tweeting jokes about these events, and as a result of resulting threats had to flee her home.[28]

In a pathbreaking 2009 article,[29] law professor Danielle Keats Citron extensively documented the scale of such attacks, particularly on women and African Americans. In the article, Citron noted evidence that such attacks, or the threat of them, significantly reduced female participation in online forums and also imposed severe privacy harms on victims. The reputation of the targets of such campaigns can also be shattered, because online threats and doxing are often accompanied by vicious falsehoods about victims – a good example being a 2007 incident in which anonymous posters used the social networking site AutoAdmit, which focused on law school admissions, to post threats and lies about a number of female law students, mainly at the Yale Law School.[30]

[27] Asia A. Eaton, Holly Jacobs, and Yanet Fuvalcaba, *Cyber Civil Rights Initiative: 2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration, A Summary Report* (June 2017), https://cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf.

[28] Caitlin Dewey, *The Only Guide to Gamergate You Will Ever Need to Read*, Washington Post (Oct. 14, 2014), www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/.

[29] Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. Rev. 61 (2009).

[30] *Ibid*. at 71–75.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

It should be noted, moreover, that the Citron article and the incidents it recounts predated or occurred in the very early years of the major social media platforms, before they exploded around 2010. Since then, social media platforms have, unsurprisingly, become a major conduit for online harassment. In particular, a 2018 report by the human rights nonprofit Amnesty International demonstrates extensively, and in deeply disturbing detail, the nature and breadth of abuse directed against women, especially prominent women, on Twitter/X.[31] The same report presents the results of an earlier online poll which demonstrates that across countries, an extraordinarily high percentage of women report abusive and misogynistic tweets directed at them, including 25 percent reporting threats of physical and sexual violence.[32] And unsurprisingly, the report excoriates Twitter/X for failing to adequately address what even Twitter/X executives concede is a huge problem.

The final, striking and unexpected feature of critiques against platforms for permitting doxing, threats, and harassment is that it has a distinct political valence. The people and institutions generally associated with such critiques, such as Professor Citron, Professor Mary Anne Franks (Citron's co-author[33] and the president of the Cyber Civil Rights Initiative, a nonprofit which serves victims of online abuse[34]), and Amnesty International, are generally associated with the political left. On what was then the political far-right, on the other hand, one finds mainly defenses of online misogyny. For example, the conservative activist Milo Yiannopoulos published a blurb at Breitbart about Gamergate titled (without irony) "Feminist Bullies Tearing the Video Game Industry Apart."[35] Breitbart, it should be remembered, is the far-right online news outlet previously led by Steve Bannon, who later ran Donald Trump's 2016 presidential campaign and then served as "senior counselor" in the first Trump White House. More broadly, there is evidence that online attacks such as the Gamergate ones were organized on online forums like 4Chan that are

[31] Amnesty International, *Toxic Twitter – A Toxic Place for Women* (March 21, 2018), www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1-1/.

[32] Amnesty International, *Toxic Twitter – Women's Experiences of Violence and Abuse on Twitter* (March 20, 2018), www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-3-2/.

[33] Danielle Keats Citron and Mary Anne Franks, *Criminalizing Revenge Porn*, 49 Wake Forest L. Rev. 345, 346 (2014); Danielle Keats Citron and Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. Chi. Legal F. 45, 69–74.

[34] *Cyber Civil Rights Initiative: CCRI Board of Directors*, https://cybercivilrights.org/about/board-of-directors/.

[35] Milo, *Feminist Bullies Tearing the Video Game Industry Apart*, Breitbart (Sept. 1, 2014), www.breitbart.com/europe/2014/09/01/lying-greedy-promiscuous-feminist-bullies-are-tearing-the-video-game-industry-apart/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

often associated with the political alt-right because of their lack of content moderation rules.[36]

None of this is to say, of course, that most conservatives support or condone the sorts of threatening and harassing behavior described in this section. Indeed, prior to 2016 few would have considered Breitbart to be a part of the mainstream right. But in an age in which Steve Bannon has served in the White House, and few on the political right speak out to condemn online harassment, a political gap has certainly emerged on this question.

2.4 HATE SPEECH

Another area in which social media platforms have faced long-standing, sharp, and consistent criticism is in their (alleged) failure to control "hate speech" on their platforms. That hate speech – which we can loosely define as attacks on individuals or groups based on characteristics such as race, sex, religion, disability, sexual orientation, and gender identity – occurs on social media platforms is of course true; and this is true despite the fact that the major platforms (including Facebook, Instagram, Twitter/X, YouTube, and TikTok) all ban hate speech. Content moderation is inevitably imperfect, and at least some critics claim (whether fairly or not) that platforms' commitment to combating hate speech is not particularly strong.[37] Indeed, in an audit commissioned by Facebook itself and published in July 2020, the company was sharply criticized for its failures in these areas.[38]

The inability or failure of platforms to fully curb hate speech, despite their written commitments to do so in their own policies, has inevitably drawn sharp attacks from the progressive left. For example, in January of 2020, a Democratic New York state senator introduced legislation that would fine platforms that failed to create adequate procedures for removing hate speech in a timely fashion.[39] In a similar vein, in August of 2020 a group of twenty state attorneys general, all Democrats, sent a joint letter to Facebook

[36] Casey Johnston, *Chat Logs Show How 4Chan Users Created #GamerGate Controversy*, Ars Technica (Sept. 9, 2014), https://arstechnica.com/gaming/2014/09/new-chat-logs-show-how-4chan-users-pushed-gamergate-into-the-national-spotlight/.

[37] 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.

[38] Facebook's Civil Rights Audit – Final Report 42–58 (July 8, 2020), https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

[39] Stacy Livingston, *New York State Senator Introduces "Social Media Hate Speech Accountability Act,"* JOLT Digest (Feb. 12, 2020), https://jolt.law.harvard.edu/digest/new-york-state-senator-introduces-social-media-hate-speech-accountability-act.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

demanding that Facebook do a better job of policing hate speech and other forms of harmful speech on its platforms.[40] And while the letter itself does not go beyond calling on Facebook to take voluntary action, in an interview with the *New York Times*, 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."[41] 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 regulation of Facebook's content moderation policies.

As a preliminary matter, it should be noted that these threats to regulate online hate speech are essentially hot air because any such regulatory efforts would be blatantly unconstitutional under the First Amendment, as interpreted by the US Supreme Court in modern times. In particular, the Court has made it clear that hate speech is fully protected under the First Amendment, absent narrow and unusual circumstances.[42] Indeed, because such speech is considered political speech on matters of public concern, it receives the very highest level of First Amendment protection.[43] And to cap things off, the Court has consistently in recent years treated efforts to suppress hate speech as almost *per se* unconstitutional viewpoint-based regulations.[44] Thus at least within the United States, any efforts to restrict online hate speech necessarily depend on the voluntary actions of platform owners rather than on the law.

Moreover, even if regulations of hate speech were constitutional, it is far from clear that they would be a good idea. Unlike in the United States, most other countries do not constitutionally protect hate speech, and many have moved to regulate it online. Perhaps the most prominent example of this is

[40] 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.

[41] *See* 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.

[42] *See* Matal v. Tam, 137 S. Ct. 1744, 1764 (2017) (plurality opinion); *ibid.* at 1766–67 (Kennedy, J., concurring in part and concurring in the judgment); R.A.V. v. City of St. Paul, 505 U.S. 377, 395–96 (1992). The narrow circumstances in which hate speech can be banned is when it constitutes a "true threat," Virginia v. Black, 538 U.S. 343, 359 (2003), or when it is directed at an individual, in person, in a way that makes the speech "fighting words," Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942). Since the fighting words doctrine is limited to in-person speech, it is of course irrelevant on the internet.

[43] *See* Snyder v. Phelps, 562 U.S. 443, 458 (2011).

[44] *See, e.g.*, *Matal*, 137 S. Ct. at 1766–67 (Kennedy, J., concurring in part and concurring in the judgment); *R.A.V.*, 505 U.S. 377, 391–92 (1992).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

Germany's NetzDG law, which became effective on January 1, 2018. NetzDG provides that 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.[45] In an attempt to comply with this law (and enforce its own Terms of Service), Facebook established a deletion center outside of Berlin staffed by over 1,200 content moderators.[46] The results, however, have been less than ideal. In particular, there have also been complaints that Facebook is, out of caution triggered by the risk of large fines, deleting legitimate posts, and that the law is chilling political speech.[47] 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.[48] The merits of a strict approach to online hate speech, therefore, remains highly disputed.

Additionally, it is far from clear that the advent of social media has increased the incidence of hate speech, in any empirically measurable way. Of course, social media has no doubt increased the *salience* of such speech, by exposing it more publicly. But it is entirely possible that the same individuals would have expressed precisely the same views in the past in private, as they do today online. Nor is it clear that the greater online salience of hate speech has any impact on general societal attitudes; to the contrary, as discussed later in more detail with respect to political polarization, what research there is suggests that polarized individuals (including hateful ones) seek out polarized content; but exposure to that content does not change preexisting beliefs.

Finally, another notable thing about critiques of online hate speech is that they too are almost entirely located in the political left. Far from encouraging greater moderation of hate speech, as noted in Chapter 1, Republican politicians regularly accuse platforms of using anti-hate speech rules to suppress conservative content. Indeed, under HB 20, the Texas legislation discussed in Chapter 1, suppression of hate speech by platforms would constitute illegal "viewpoint-based" content moderation (since, as noted earlier, under

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

[46] Katrin Bennhold, *Germany Acts to Tame Facebook, Learning from Its Own History of Hate*, N.Y. Times (May 19, 2018, 10:45 AM), www.nytimes.com/2018/05/19/technology/facebook-deletion-center-germany.html.

[47] Rebecca Zipursky, Note, *Nuts about NETZ: The Network Enforcement Act and Freedom of Expression*, 42 Fordham Int'l L.J. 1325, 1359–60 (2019); *see* 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/.

[48] *See* Rebecca Zipursky, Note, *Nuts about NETZ: The Network Enforcement Act and Freedom of Expression*, 42 Fordham Int'l L.J. 1325, 1360–62 (2019).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

established Supreme Court precedent hate speech constitutes a viewpoint). In short, hate speech is yet another area where the political left urges more content moderation, and the political right urges less.

2.5 FILTER BUBBLES, IDEOLOGICAL SILOS, AND POLARIZATION

A final, important critique of social media platforms, and indeed of the internet more generally, is that online speech has contributed to ever-increasing political polarization, both in the United States and around the world. As early as 2001, law professor Cass Sunstein, then at the University of Chicago (where I studied with him) and now at Harvard, published a seminal book titled *Republic.com* which expressed concerns of this nature.[49] Sunstein's basic argument, or perhaps more accurately his worry, was that as speech moves onto the internet (this was well before social media became important), individuals would begin filtering the news and speech they were exposed to, limiting it in ways that would confirm their own preexisting beliefs and biases. The consequence would be a society fragmented into groups of like-minded citizens, with little or no communication across those groups. In other words, his concern was that the internet would increase political tribalism and polarization, perhaps to an unsustainable level.

When Sunstein made this argument in 2001, many people (including myself) were skeptical. After all, the impact of the internet was to make speech cheap, and so to enable ever more speech by ever more people. Wouldn't this lead to people being exposed to more perspectives than before, when the institutional media dominated public discourse? Whatever the truth of the matter in 2001, the advent and growth of social media platforms, and their eventual dominance of online discourse, made many of those on the political left (where Sunstein himself firmly belongs – he served as the so-called Regulatory Czar in the Obama Administration) come to see Sunstein's arguments as prophetic. And indeed, in 2017 Sunstein published another book, *#Republic*, which updated his argument for the social media era.[50]

There seems little doubt that there is a large degree of ideological conformity in the content to which social media users are exposed. In other words, social media use really does tend to confirm existing biases and beliefs, just as Sunstein predicted. The reasons for this phenomenon are complex. Many social media critics argue that the problem lies in the algorithms that

[49] Cass R. Sunstein, *Republic.com* (2001).
[50] Cass R. Sunstein, *#Republic* (2017).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

platforms use to decide what content to display to individual users. Because platforms care first and foremost about maximizing engagement (to maximize profits), these algorithms choose content that matches users' own previously expressed interests and views. The result is confirmation of those beliefs over time, leading to increasingly firmly held views. And, of course, because different individuals have different beliefs confirmed and strengthened, the broader social result is polarization.

Of course, no serious person would claim that the internet, or social media, are the only or even primary causes of increasing political polarization in the United States, which after all long predates these technologies. But important voices on the left, including in a report published by the Brookings Institute (perhaps the epicenter of progressive thinking), endorse the position that platform algorithms play an important role in fueling the phenomenon.[51] Jonathan Haidt, a highly influential scholar at New York University has also advocated this position,[52] as of course have others.[53] Haidt's argument in particular focuses not just on how platform algorithms select content, but more on ways in which platforms are designed to encourage extremist content because such content is much more likely to be retweeted, shared, or liked than more nuanced views. But ultimately, the result is the same, which is that the nature of the operation of the platforms creates exposure to content that confirms, and exaggerates, preexisting beliefs. And the common theme of all of these arguments is to link the desire of platforms to maximize engagement to the phenomenon of polarization.

There is, however, a somewhat peculiar aspect to this criticism of social media. Critics claim that social media recommends divisive content to users because such content is seen to maximize user engagement – and user engagement is, after all, the ultimate goal of platforms who make money by selling targeted advertising. But it is important to notice something. When critics complain that social media firms maximize engagement through their prioritization algorithms, in plain English what they are saying is that social media is at fault for emphasizing content that users *like*, that they want to see. But one might ask,

[51] Paul Barrett, Justin Hendrix and Grant Sims, *How Tech Platforms Fuel U.S. Political Polarization and What Government Can Do about It*, Brookings (Sept. 27, 2021), www.brookings.edu/blog/techtank/2021/09/27/how-tech-platforms-fuel-u-s-political-polarization-and-what-government-can-do-about-it/.

[52] 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/.

[53] Bill Whitaker, *Social Media's Role in America's Polarized Political Climate*, CBS News (Nov. 6, 2022), www.cbsnews.com/news/social-media-political-polarization-60-minutes-2022-11-06/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

is that not the job of any entity that provides entertainment? And one might also ask where the real responsibility for bad social outcomes lies here – in social media companies that feed their users' desires or in the users themselves?

It should be noted, moreover, that not everyone (on the political left or otherwise) agrees that platform algorithms are the main reason why social media increases polarization. An important counterpoint to those who point to algorithms is that the "blame" for platforms increasing polarization falls not on the platforms but rather on us, the users. In particular, law professor Jane Bambauer and colleagues at the University of Arizona argue that the real culprit is the internet's enabling of "cheap friendship," which permits people to socialize online almost exclusively with like-minded people – something that is much harder to do in the physical world.[54] And it is this tendency of users to cluster with those like them that produces the confirmation of beliefs and biases, including on seemingly factual matters, which fuels polarization. It should be noted, however, that whether the driving force is algorithms or friend selection, these arguments all support the view that using social media significantly increases polarization.

Of course, not everyone agrees that social media *does* meaningfully influence polarization. Unsurprisingly, senior figures at Meta, the owner of Facebook and Instagram, continue to dispute this point.[55] And importantly, recent research tends to support this view. In particular, a carefully designed recent study published in the leading scientific journal *Nature* suggests that while social media platforms do indeed have a tendency to present users with content consistent with their preexisting beliefs, that exposure does *not* tend to increase political polarization.[56] Of course, there is a correlation between political polarization and politically oriented social media use; but all that shows is that highly polarized individuals (unsurprisingly) tend to seek out the highly politicized corners of social media and the internet. But as with Professor Bambauer's analysis, this is consistent with the thesis that the problem is not platforms and their algorithms; the problem is on the demand side, situated squarely in individual users.

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

[55] *Mark Zuckerberg Opening Statement Transcript: House Hearing on Misinformation* (March 25, 2021), www.rev.com/blog/transcripts/mark-zuckerberg-opening-statement-transcript-house-hearing-on-misinformation; Nick Clegg, *You and the Algorithm: It Takes Two to Tango* (March 31, 2021), https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2.

[56] Brendan Nyhan, Jaime Settle, Emily Thorson, et al., *Like-Minded Sources on Facebook Are Prevalent but Not Polarizing*, Nature (July 27, 2023), www.nature.com/articles/s41586-023-06297-w.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

Indeed, another study conducted by many of the same authors and published in *Science*, the other leading scientific journal, strongly suggests that, contrary to what many critics claim, features such as resharing buttons on social media also do not alter beliefs or accelerate polarization. To the contrary, and perversely, the study suggests (though its findings in this regard are less definitive) that disabling such features decreases users' knowledge about current news, without any accompanying benefits.[57] In other words empirical science, that darling of the political left, suggests that progressive palpitations about the polarizing effects of social media might well be baseless.

In truth, it is almost impossible to separate out the impact of the spread of social media from other societal trends contributing to political polarization in the United States, such as increasing racial diversity and the economic impact of globalization on working class Americans. And it is of course possible that social media has played some role in increasing political polarization – though again, the data supporting this view is largely absent. But the left and center's tendency to attribute political polarization mainly to social media, or even the internet, while largely neglecting other causes, such as public policies and public rhetoric associated with the left, is either perplexing or breathtakingly cynical.

Finally, it should be noted that these concerns about polarization *are* of the center and left, and do not appear to be shared by more conservative leaders and thinkers. To the contrary, prominent conservative politicians, such as Marjorie Taylor Greene, President Trump, and Governor Ron DeSantis of Florida, seem to have adopted a strategy of stoking political polarization for personal gain. Nor is this a new, or online, phenomenon, as illustrated by then Speaker of the House Newt Gingrich's conduct during the 1990s. And of course there are plenty of political figures on the far left that behave similarly. But it remains the case that concerns about the role of social media in stoking political polarization and extremism have for the past decade been, and remain, a mainstay of the center-left in the United States.

2.6 CHILDREN, ADDICTION, AND BODY IMAGE

Finally, we should consider one of the most potent recent attacks on social media – this time one shared by the political left and right. It is the argument that social media is causing intense harm to the mental health of children,

[57] Andrew M. Guess, Neil Malhotra, Jennifer Pan, et al., *Reshares on Social Media Amplify Political News but Do Not Detectably Affect Beliefs or Opinions*, SCIENCE (July 27, 2023), www.science.org/doi/full/10.1126/science.add8424.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

including by promoting addictive behavior, by contributing to elevated levels of depression and anxiety, and by promoting body image issues among teenagers (especially teenage girls). These concerns have generated a bestselling book by Jonathan Haidt,[58] a proposal by Republican Senator Josh Hawley of Missouri to prohibit social media platforms from offering accounts to children under the age of 16,[59] and, that infallible indicator of the concerns of the intelligentsia, an article in *The Economist*.[60] Given children's perceived heightened vulnerability to undue influences, these concerns unsurprisingly have received a great deal of attention and have led both the European Union and California to adopt legislation designed to protect children.[61]

Before jumping to extreme solutions (such as Senator Hawley's proposal to ban social media for most children), however, it is important to separate, examine, and dissect these concerns more carefully. Take, for example, the most sweeping (and vague) concern, that the combination of the internet, social media, and smart phones has created an addiction crisis among children, interfering with their socialization. The truth is, though, that while everyone is familiar with the phenomenon of "doom scrolling" (something hardly limited to children), there is no good, accepted definition of what internet or social media "addiction" even is.[62] And insofar as casual commentators are associating "addiction" with screen time on smart phones, the difficulty is that some of the most common uses of smart phones by minors, such as texting, are social activities which have merely displaced things such as phone calls. In my long-distant childhood, in an age before even cell phones, parents regularly chastised teenagers for spending too much time on the telephone, but no one in that era pulled out the term "addiction."

Turning from amorphous concepts of addiction, however, to more measurable mental health concerns such as depression and suicide, some serious

[58] Jonathan Haidt, The Anxious Generation: How the Great Rewiring of Childhood is Causing an Epidemic of Mental Illness (2024).

[59] *New: Hawley Introduces Two Bills to Protect Kids Online, Fight Back against Big Tech*, Josh Hawley: U.S. Senator for Missouri (Feb. 14, 2023), www.hawley.senate.gov/new-hawley-introduces-two-bills-protect-kids-online-fight-back-against-big-tech/.

[60] *Time for a Digital Detox?: Demands Grow to Restrict Young People's Access to Phones and Social Media*, The Economist, April 20, 2024, at 87.

[61] *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 (Digital Services Act)* Art. 28(2), https://eur-lex.europa.eu/eli/reg/2022/2065/oj; European Commission, Directorate-General for Communications Networks, Content and Technology, *The Digital Services Act (DSA) Explained – Measures to Protect Children and Young People Online*, Publications Office of the European Union, 2023, https://data.europa.eu/doi/10.2759/576008; California Age-Appropriate Design Code Act, Cal. Civ. Code §§ 1798.99.28–1798.99.40.

[62] Caroline Miller, *Is Internet Addiction Real?*, Child Mind Institute (Dec. 8, 2023), https://childmind.org/article/is-internet-addiction-real/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

questions undoubtedly exist. There is no doubt that during the period when social media use has become prevalent (starting around 2010), rates of depression among adolescents have increased sharply, especially among teenage girls.[63] During that same period of time, suicide rates among young people aged 12–17 in the United States increased by almost 50 percent (suicide rates increased across almost all age groups, but the most among the young).[64] And international data suggests that from around 2013 to 2021, suicide rates among adolescent girls (aged 10–19) increased by 50 percent in developed countries (though it should be noted that that group also has by far the lowest suicide rate of any demographic – boys/men have substantially higher suicide rates than girls/women at every age, and suicide rates increase consistently with age).[65] Given the obvious correlation in time between these objective indicia of an adolescent mental health crisis and the rise of social media, it is very tempting to conclude (as many have) that there is a direct causal connection. And perhaps such a causal connection does indeed exist.

The difficulty, however, is that the proof of the pudding is in the eating, and on this issue there simply is no proof. There have been an enormous number of studies conducted in recent years examining the link between social media use and mental health, and politicians such as Senator Hawley, as well as some journalists, have seized upon some of them as a call to arms to regulate social media use by adolescents. In 2024, however, a blue ribbon committee of experts convened by the National Academies of Sciences, Engineering, and Medicine issued a report on the topic, *Social Media and Adolescent Health*, which painstakingly reviewed the extant literature. Its strikingly straightforward conclusion, stated in the summary, was that "[t]he committee's review of the literature did not support the conclusion that social media causes changes in adolescent health at the population level."[66] And on the specific topic of adolescent depression, the report notes that "[s]tudies looking at the association between social media use and feelings of sadness over time have largely found small to no effects."[67] Of course, this conclusion does not rule out the possibility that social media use can negatively affect the mental health of some

[63] Sylvia Wilson and Nathalie M. Dumornay, *Rising Rates or Adolescent Depression in the United States: Challenges and Opportunities in the 2020s*, 70 J. Adolesc. Health 354, 354–55 (2022), www.jahonline.org/article/S1054-139X(21)00646-7/fulltext.

[64] Heather Saunders and Nirmita Panchal, *A Look at the Latest Suicide Data and Change Over the Last Decade*, KFF (Aug. 4, 2023), www.kff.org/mental-health/issue-brief/a-look-at-the-latest-suicide-data-and-change-over-the-last-decade/.

[65] *Time for a Digital Detox?*, *supra* n. 60.

[66] Nat'l Academies of Sciences, Eng'g, and Med., Social Media and Adolescent Health 5 (2024) (henceforth "*Academies' Report*").

[67] *Ibid*. at 104.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

individuals, adolescent or otherwise (or for that matter, improve individuals' mental health). But it does strongly suggest that the proposition that social media use is the cause of a societal mental health crisis among adolescents remains unproven.

Digging down deeper into the Academies' report, the reasons for this failure of proof become clearer. The most fundamental problem is the difficulty in teasing out causation. Even if the data shows a correlation between social media use and mental health issues (and the data suggests that a small such association may exist), it is impossible to say whether the reason for this is that social media use harms mental health or, equally plausibly, that adolescents with mental health challenges are more likely to turn to social media.[68] Another problem is that the period when the relevant research was conducted overlaps with the extreme political divisiveness of the 2016 election and Trump presidency, culminating of course with the COVID-19 pandemic and accompanying lockdowns. Especially during the latter period, both social media use and mental health challenges very predictably increased among adolescents (and adults), but teasing out causation here is essentially impossible. Finally, the sheer variety of platforms encompassed by the term "social media," from Instagram to Twitter/X to chats between online gamers, means that studies focused on social media use generally are of limited value.

But wait, I imagine many of you are thinking, didn't the Surgeon General of the United States issue an Advisory regarding social media use by children and adolescents? Indeed he (Surgeon General Vivek Murthy) did, in 2023.[69] And based on that Advisory, the Surgeon General recommended to Congress that warning labels for adolescents be placed on social media platforms (as well as writing an op-ed in the *New York Times* about warning labels).[70] So doesn't that establish that there is solid scientific evidence that social media causes harm, just as the Surgeon General's mandatory labels for tobacco products were based on solid scientific evidence that smoking causes lung cancer? Not exactly.

If one reads the Surgeon General's Advisory through, one sees many studies of the sort discussed in the Academies' report that find some correlation between social media use and depression.[71] But for the reasons just discussed,

[68] *Ibid*. at 93–94.

[69] *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*, U.S. Public Health Service (2023), www.hhs.gov/surgeongeneral/priorities/youth-mental-health/social-media/index.html.

[70] Vivek H. Murthy, *Surgeon General: Why I'm Calling for a Warning Label on Social Media Platforms*, N.Y. Times (June 17, 2024), www.nytimes.com/2024/06/17/opinion/social-media-health-warning.html.

[71] *Social Media and Youth Mental Health*, *supra* n. 69, at 7–8.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

those studies do *not* demonstrate a causal connection between social media use and mental health challenges. Indeed, the Surgeon General's Advisory is phrased very carefully, to reflect this; the key sentence reads as follows: "At this time, we do not yet have enough evidence to determine if social media is sufficiently safe for children and adolescents."[72] In other words, we don't know. And toward the end, the Advisory analogizes social media to approval of new medications by the FDA, where approval is contingent on *proof* that the medication is safe and effective.[73]

But this is a category error. It is indeed true that with prescription medications, the United States does not permit their sale unless the relevant pharmaceutical company can prove their safety. And, as the Advisory notes, that is the same approach we take to children's toys (the regulating agency there is the Consumer Product Safety Commission).[74] But that is *not* the approach we in this country take to free speech – which, after all, is what social media is. To the contrary, such an approach to speech, approval before use, is in the speech context called a "prior restraint," and it is almost automatically unconstitutional under the First Amendment to the US Constitution.[75] Under our constitutional regime, if the government wants to restrict speech, the burden is on the *government* to prove that the speech at issue causes significant harm. Moreover, the Supreme Court has made it clear that the government bears the same burden in restricting speech directed at children as it does more generally.[76] And it is quite clear that absent further, more definitive research, that burden has not been met.

Finally, let us close with one of the most often-expressed, specific concerns about social media use by adolescents, which is that it greatly increases body image issues and related eating disorders among adolescent girls (similar concerns regarding boys have not been much studied). There are indeed some studies suggesting a connection between (some forms of) social media use and body image issues. Indeed, the release to Congress and the media in 2021, by former Facebook employee Francine Haugen, of an internal study by Facebook arguably demonstrating such an effect with respect to Instagram use was at the center of the ensuing "Facebook Files" scandal (recall that Meta, then Facebook, owns Instagram).[77] And it is certainly plausible that constant

[72] *Ibid*. at 4.
[73] *Ibid*. at 14.
[74] *Ibid*.
[75] New York Times Co. v. U.S., 403 U.S. 713, 714 (1971) (per curiam).
[76] Brown v. Entertainment Merchants Ass'n, 564 U.S. 786, 799 (2011).
[77] Georgia Wells, Jeff Horwitz, and Deepa Seetharam, *The Facebook Files: Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, Wall Street Journal (Sept. 14, 2021), www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

exposure to a platform such as Instagram, which focuses on photographs of peers, often filtered or altered ones, *could* contribute to body image issues and eating disorders.

But once again, the full truth is more complicated. As the Academies' report notes, even in this area the usual causation problems remain. As the report states, concerns about media depictions of female beauty driving body image issues long predates social media.[78] Furthermore, the report notes, while social media use certainly *might* contribute to body image issues and eating disorders, "the psychological factors that influence the development of eating disorders … can also manifest in disordered behaviors such as overuse of social media."[79] In other words, the causation might well run in the opposite direction, with disorders leading to social media use rather than vice versa. But all that said, there concededly is a realistic risk that certain forms of social media use contribute to the existing, exceedingly concerning problem of poor self-images among adolescent girls.

But even if that point *is* conceded, two complications must be addressed. The first is that while critics often attribute this problem to "social media," in fact the problem appears to be almost entirely associated with specific, photograph-focused platforms, most of all Instagram and Snapchat (though oddly, only the former gets media attention – presumably because of the media's joy in bashing "Big Tech"). After all, while exposure to political trolling on Twitter/X may well be bad for mental health in a colloquial sense, it is hardly likely to contribute to body image issues. As such, using this argument to justify limiting adolescent access to *all* social media is misguided and vastly overbroad.

Furthermore, as law professor Eric Goldman has pointed out, there is a more fundamental complication here that the critics ignore. Goldman focuses on the October 2019 internal Facebook study regarding Instagram, the release of which in 2021 by Francine Haugen triggered a firestorm. The headline chart in this study has been cited, fairly, for the proposition that Instagram use makes almost 20 percent of US teens, and over 20 percent of US teen girls, feel worse about themselves (the number goes up to 25 percent for teen girls in the UK).[80] But what the critics fail to mention is that 41 percent of US respondents reported that Instagram had no effect on their self-worth, and that another 41 percent (including 37 percent of US

[78] *Academies' Report*, *supra* n. 66, at 97.

[79] *Ibid.*

[80] Eric Goldman, *The "Segregate-and-Suppress" Approach to Regulating Child Online Safety*, at 21, Working Paper on File with Author (July 30, 2024); *see also* Wells, Horwitz, and Seetharam, *supra* n. 77.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

girls) reported that exposure to Instagram made them feel *better* about themselves. Even among teen girls in the UK, the demographic whose study data was most troubling, 30 percent of respondents reported that Instagram made them feel better about themselves.[81] In other words, even among adolescent girls, Instagram apparently makes substantially more of them feel better about themselves than worse (the numbers are much more positively skewed for boys). So, if we take *all* of this data seriously, rather than cherry-picking, blocking adolescent access to Instagram will presumably make more teenagers feel worse about themselves than better. What is an honest, objective observer to make of that?

None of this is to say that concerns about links between children's use of social media and mental health issues are unreasonable or fanciful. They are not; indeed, they are perfectly reasonable and widely shared. But they are not proven, and the underlying dynamic is far more complex than critics acknowledge. In that world, it would certainly be perfectly reasonable for concerned parents to monitor or limit their children's use of social media. Furthermore, there are obvious and largely unrelated reasons that might justify school policies restricting smart phone use in classrooms (the obvious one being distraction). But when we come to regulatory interventions, skepticism seems in order.

2.7 IMPLICATIONS AND CONTRADICTIONS

There are several points to notice about progressive critiques of social media. The first, and perhaps most important, is that other than amorphous concerns about adolescent mental health, the criticisms ultimately come down to a claim that social media should suppress more content than it does because the content is socially harmful. This point is obviously true about the various attacks on the spread of disinformation, hate speech, or harassing speech; but it also is true of the polarization claim, which ultimately comes down to a demand to suppress or hide inflammatory content on platforms. In other words, progressives want platforms to act as gatekeepers of information, shielding the public from content deemed (by whom is not exactly clear) to be harmful. In Chapter 5, I will explain why I think this is a very bad idea, but for now it is the nature of the claim that matters.

Secondly, many or most progressive critiques, like conservatives claims of platform bias, stand on shaky empirical grounds. Why that is so with respect to hate speech and political polarization has already been discussed. But as

[81] Goldman, *supra* n. 80, at 21–22.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

it turns out, the same is true with respect to mis- and disinformation. In an extremely thoughtful article in the *New Yorker*, Professor Manvir Singh of the University of California at Davis (my own institution) summarizes a great deal of empirical research that raises serious doubts about whether exposure to false information actually changes people's actions.[82] Professor Singh summarizes a swath of social science research which supports the view that people have two distinct kinds of beliefs: "factual" beliefs, rooted in data about the real world, and "symbolic" beliefs that are more akin to faith about the abstract nature of the world. Factual beliefs, the evidence shows, are susceptible to being changed by exposure to evidence; but symbolic beliefs by and large are not, because they are driven more by social ends such as group solidarity and the reinforcement of political identity.

Crucially, however, what the studies Singh summarizes tend to show is that individuals are far more likely to *act* based on their factual beliefs than their symbolic ones, if those actions have consequences for themselves. In other words, when individuals have "skin in the game," they actually do care about empirical facts. It should be emphasized that this does not mean that people's symbolic beliefs are not authentic – to the contrary, they generally seem to be. But individuals clearly recognize, at some level, that symbolic beliefs are different in kind from factual ones, leading both to a greater willingness to reassess factual beliefs than symbolic ones, and a greater willingness to act upon them. All of which does not suggest that mis- and disinformation is not abundant on social media – of course it is. But it does suggest that the real-world impacts of such content are quite limited. And that in turn suggests that the apocalyptic fears that Singh also gathers, suggesting that the spread of online mis- and disinformation spells the end of democracy and of social cohesion, are greatly overstated if not a form of mass hysteria.

Furthermore, it is also deeply unclear, as progressive critics tend to assume, that it is the architecture and algorithms deployed by social media platforms that are causing the spread of false information. To the contrary, a compelling recent paper argues that the online spread of what it calls "bullshit" is not a product of platform design but rather of consumer demand. In other words, the fault lies not in platforms but in ourselves. To this point, the paper argues that rather than platforms favoring low quality content, it is individuals that seek it out (which admittedly incentivizes platforms to serve up such content to those individuals). But comfortingly, the paper also argues (citing empirical

[82] Manvir Singh, *Don't Believe What They're Telling You about Misinformation*, New Yorker (April 15, 2024), www.newyorker.com/magazine/2024/04/22/dont-believe-what-theyre-telling-you-about-misinformation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

evidence) that the actual market for bullshit is quite limited, focused on a small percentage of users who start out with highly polarized beliefs.[83] All of which again suggests that the progressive attack on platforms as the source of false information is misguided (and that any solutions targeted at platforms, such as requiring changes to their algorithms, will be largely ineffective).

A third point about progressive critiques is that the social ills which they attribute to social media, in fact, are hardly limited to the social media ecosystem and long predate the spread of online platforms. Bigotry, hate speech, misogyny and threats of gender-based violence, conspiracy theories, and irrational attitudes toward science, including fear of vaccines, have all, unfortunately, been pervasive elements of the national culture of the United States for many decades, if not throughout our history. On conspiracy theories in particular, the historian Richard J. Hofstadter noted the influence of such thinking on the American political right in *The Paranoid Style in American Politics*, an essay published in 1964 (in *Harper's Magazine*, and later republished as a book)[84] that in turn was based on a 1959 lecture. And the historical pervasiveness of racism and racist speech in the United States is hardly in need of proof – though the *New York Times*'s 1619 Project does admirable work on that issue.[85]

Of course, this is not to say that the internet in general, and social media platforms in particular, have not increased the breadth and impact of such thinking. Perhaps QAnon has had a greater impact on the conservative movement today than the John Birch Society did in the 1960s, thanks to social media. Perhaps because of social media, vaccine skepticism during the COVID-19 pandemic was more pervasive than in earlier times and had greater health impacts. But then again, perhaps not. There are many trends in American society and politics, other than or in addition to new communications technology, that might explain the disappointingly low quality of public discourse in the modern era. Fox News, after all, predates social media by a decade and has almost certainly contributed more to polarization than the internet.[86]

[83] Lia Greenberg, Katherine Marin, Jessica Sparks, and Jane Bambauer, *The Demand for Bullshit, in* The Elgar Companion to Freedom of Speech and Expression (Ashutosh Bhagwat and Alan Chen eds., in press).

[84] Richard J. Hofstadter, The Paranoid Style in American Politics and Other Essays (1964).

[85] *The 1619 Project*, N.Y. Times, www.nytimes.com/interactive/2019/08/14/magazine/1619-america-slavery.html.

[86] Gregory J. Martin and Ali Yurukoglu, *Bias in Cable News: Persuasion and Polarization*, 107 Am. Econ. Rev. 2565 (2017); David E. Brockman and Joshua L. Kalla, *Selective Exposure and Partisan Echo Chambers in Television News Consumption: Evidence from Linked Viewership, Administrative, and Survey Data*, Working Paper (April 17, 2023), https://osf.io/b54sx/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003

Finally, it is noteworthy how stridently progressive critics of social media believe, and insist, that social media firms should have moral, ethical, and (eventually) legal obligations to alleviate or cure the social ills associated with allegedly harmful content. Yet no one appears inclined to impose such obligations on other forms of media, communications technologies, or industries. As an example, consider the fact that Instagram is regularly attacked for exacerbating body image issues among teenage girls and is held responsible for these impacts in public discourse.[87] Yet obviously body imagine issues among teenage girls did not suddenly arise in 2009 (when Instagram was launched); Mark Zuckerberg, in short, has not destroyed an idyllic past.[88] To the contrary, our popular culture has pushed negative body images, especially onto teenage girls, for many, many decades. And yet, there is a remarkable lack of pressure to impose similar ethical or legal obligations as those urged for social media onto Hollywood, fashion magazines, or for that matter the fashion industry. Similarly, while as noted earlier Fox News's role in exacerbating political polarization is well accepted, no one suggests regulating Fox News to limit divisive content. The inconsistency is striking.

[87] *See, e.g.*, Billy Perrigo, *Instagram Makes Teen Girls Hate Themselves. Is That a Bug or a Feature?*, TIME (Sept. 16, 2021), https://time.com/6098771/instagram-body-image-teen-girls/; Catherine Pearson, *Alarming New Report Shows Just How Toxic Instagram Is for Body Image*, HUFFPOST (Oct. 4, 2021), www.huffpost.com/entry/new-report-instagram-body-image_l_615b1419e4b008640eb738e5.

[88] For an unusual perspective making this point *see* Jessica Grose, *The Messy Truth about Teen Girls and Instagram: You Can't Blame Social Media for Everything*, N.Y. TIMES (Oct. 13, 2021), www.nytimes.com/2021/10/13/parenting/instagram-teen-girls-body-image.html.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 16:08:22, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009547703.003