

---

## The Politics and Perverse Effects of the Fight Against Online Medical Misinformation

*Evelyn Douek*

**ABSTRACT.** Social media platforms’ moderation of medical misinformation has emerged as one of the most contentious political issues of our time. This Essay traces the evolution of platforms’ approach to medical misinformation during the COVID-19 pandemic and argues that the politicization of platforms’ actions is a product, at least in part, of the difficulty of defining “medical misinformation” as a coherent category. Going forward, platforms’ approach to moderating medical speech should reflect a principled view of the role they can, and cannot, play as gatekeepers of medical knowledge, rather than shifting with the political winds. That role, this Essay argues, and as First Amendment case law suggests, needs to be informed by the institutional characteristics of platforms as social spaces, and cannot assume—as platforms seemed to do at the start of the pandemic—that there is a sharp divide between medical and political speech.

### INTRODUCTION

Social media platforms’ moderation of medical misinformation has emerged as one of the biggest political controversies of our time. It has been the focus of congressional committees,<sup>1</sup> a Supreme Court case,<sup>2</sup> and countless newspaper headlines.<sup>3</sup> It was also a prominent theme of the 2024 presidential race. When

---

1. Jared Gans, *These Republicans Will Serve on Panels to Probe COVID-19, “Weaponization” of Government*, HILL (Jan. 24, 2023, 7:19 PM ET), <https://thehill.com/homenews/house/3829146-these-republicans-will-serve-on-panels-to-probe-covid-19-weaponization-of-government> [<https://perma.cc/7D8B-LH5J>].

2. *Murthy v. Missouri*, 144 S. Ct. 1972 (2024).

3. See, e.g., Philip Hamburger, *Biden’s Social-Media Censorship Harms Us All*, WALL ST. J. (Sept. 26, 2023, 6:03 PM ET), <https://www.wsj.com/articles/bidens-social-media-censorship-harms-us-all-6cec6e55> [<https://perma.cc/3E7A-PWBC>]; Andy Hoffman & Clara Hernanz Lizarraga, *WHO Warns About the Dangers of Health Misinformation as Musk Takes over Twitter*,

then-candidate J.D. Vance was asked during the vice-presidential debate the (for him, difficult) question of whether Donald Trump had lost the 2020 election, he answered by pivoting to what he argued was the more important issue going forward—“Did Kamala Harris censor Americans from speaking their mind in the wake of the 2020 COVID situation?”<sup>4</sup>

Vance’s seeming non sequitur was not entirely surprising. Outrage at the heavy-handed way that social media platforms moderated COVID-19 related content also informed President-elect Donald Trump’s political rhetoric during his campaign<sup>5</sup> and now looks as if it will be a core plank of his administration’s policy agenda. As of this writing, President-elect Trump is set to name a number of people to high-level roles in his administration who have been vocal in decrying what they have cast as an unprecedented censorship campaign by social media platforms during the COVID-19 pandemic.<sup>6</sup> Trump’s pick to lead the Department of Health and Human Services, Robert F. Kennedy, Jr., has himself spread falsehoods and conspiracy theories on a wide variety of health-related issues, most prominently vaccines,<sup>7</sup> and has been party to multiple (unsuccessful) lawsuits alleging that platforms’ restrictions on such posts violate his and others’ First Amendment rights.<sup>8</sup> Trump’s proposed nominee to head the National Institutes for Health, Dr. Jay Bhattacharya, rose to prominence as a critic of the public-health response to the pandemic<sup>9</sup> and went all the way to the Supreme

---

BLOOMBERG (Apr. 26, 2022, 12:37 PM CDT), <https://www.bloomberg.com/news/articles/2022-04-26/musk-twitter-approach-spurs-who-warning-on-health-misinformation> [<https://perma.cc/7CV5-5RSE>].

4. Nia Prater, *J.D. Vance Blasted for Non-Answer on January 6 in VP Debate*, N.Y. MAG.: INTELLIGENCER (Oct. 2, 2024), <https://nymag.com/intelligencer/article/vance-january-6-vp-debate.html> [<https://perma.cc/62YA-NEFH>].
5. Emily Birnbaum, *Donald Trump Targets Social-Media ‘Censorship’ for 2024 Campaign*, BLOOMBERG (Dec. 15, 2022, 1:48 PM CST), <https://www.bloomberg.com/news/articles/2022-12-15/trump-targets-social-media-censorship-in-2024-themed-unveiling> [<https://perma.cc/RG4Q-9PY4>].
6. Benjamin Mazer, *Revenge of the COVID Contrarians*, ATLANTIC (Nov. 25, 2024), <https://www.theatlantic.com/health/archive/2024/11/covid-revenge-administration/680790> [<https://perma.cc/ZNA5-J4BM>].
7. Zeynep Tufekci, *How Robert F. Kennedy Jr. Could Destroy One of Civilization’s Best Achievements*, N.Y. TIMES (Nov. 16, 2024), <https://www.nytimes.com/2024/11/16/opinion/rfk-jr-health-secretary-trump.html> [<https://perma.cc/46TV-YD6T>].
8. *Children’s Health Def. v. Meta Platforms, Inc.*, 112 F.4th 742 (9th Cir. 2024); *Kennedy v. Google LLC*, No. 23-3411, 2024 WL 3934326 (9th Cir. Aug. 26, 2024); *Kennedy v. Meta Platforms, Inc.*, No. 24-CV-02869, 2024 WL 4031486 (N.D. Cal. Sept. 3, 2024); *Kennedy v. Biden*, No. 23-CV-00381, 2024 WL 3879510 (W.D. La. Aug. 20, 2024); *Kennedy v. Warren*, 66 F.4th 1199 (9th Cir. 2023).
9. Dan Diamond, *Jay Bhattacharya, an NIH Critic, Emerges as a Top Candidate to Lead the Agency*, WASH. POST (Nov. 17, 2024), <https://www.washingtonpost.com/health/2024/11/16/nih->

Court to seek redress for restrictions placed on his social media accounts when he posted views contrary to the medical establishment.<sup>10</sup> Meanwhile, Trump's favored Chair of the Federal Communications Commission has vowed to take on "the censorship cartel" of tech companies who "silenced Americans for doing nothing more than exercising their First Amendment rights."<sup>11</sup> The list goes on.<sup>12</sup> Backlash against the moderation of medical misinformation is thus almost certain to remain a key theme in our politics for the foreseeable future.

The regulation of medical misinformation was not always so politically loaded. At the start of the COVID-19 pandemic, platforms were widely praised for more aggressively moderating false information about the new and scary virus and were lauded for finally becoming more responsible custodians of the public sphere.<sup>13</sup> Now, less than half a decade later, platforms are trying to distance themselves from this approach. At least one social media platform has prominently rolled back its policies against COVID-19 misinformation,<sup>14</sup> and another has said that during the pandemic it "made some choices that . . . [it] wouldn't make today."<sup>15</sup> Platforms clearly believe that policing medical

---

director-jay-bhattacharya-covid-great-barrington-declaration [https://perma.cc/C97W-5TSK].

10. *Murthy v. Missouri*, 144 S. Ct. 1972, 1989 (2024).
11. Brendan Carr (@BrendanCarrFCC), X (formerly TWITTER) (Nov. 15, 2024, 8:45 AM ET), <https://x.com/BrendanCarrFCC/status/1857419658812440927> [https://perma.cc/8FUG-UBXZ].
12. See, e.g., Noah Weiland, Margot Sanger-Katz & Dani Blum, *Trump Plans to Nominate Dr. Mehmet Oz to Oversee Medicare, Medicaid and Obamacare*, N.Y. TIMES (Nov. 22, 2024), <https://www.nytimes.com/2024/11/19/us/politics/trump-dr-oz-medicare-medicaid.html> [https://perma.cc/F483-PNVL].
13. Evelyn Douek, *The Year that Changed the Internet*, ATLANTIC (Dec. 28, 2020), <https://www.theatlantic.com/ideas/archive/2020/12/how-2020-forced-facebook-and-twitter-step/617493> [https://perma.cc/Q8M7-LFZN]; Steven Levy, *Has the Coronavirus Killed the Techlash?*, WIRED (Mar. 20, 2020, 9:00 AM), <https://www.wired.com/story/plaintext-has-the-coronavirus-killed-the-techlash> [https://perma.cc/76FV-MRJQ].
14. *Twitter Rolls Back COVID Misinformation Policy*, REUTERS (Nov. 29, 2022, 4:11 PM EST), <https://www.reuters.com/technology/twitter-rolls-back-covid-misinformation-policy-2022-11-29> [https://perma.cc/259H-884P].
15. Letter from Mark Zuckerberg, CEO, Meta Platforms, Inc., to Rep. Jim Jordan, Chairman of H. Comm. on the Judiciary (Aug. 26, 2024), <https://prod-i.a.dj.com/public/resources/documents/meta-letter-o82024.pdf> [https://perma.cc/VL8P-SMWH]; see @JudiciaryGOP, X (formerly TWITTER) (Aug. 26, 2024), <https://x.com/JudiciaryGOP/status/1828201780544504064> [https://perma.cc/5425-QGK9]; see also Alex Heath, *Meta Says It's Mistakenly Moderating Too Much*, VERGE (Dec. 3, 2024, 8:00 AM EST), <https://www.theverge.com/2024/12/3/24311513/meta-content-moderation-mistakes-nick-clegg> [https://perma.cc/L6D3-YR5Y] (quoting Meta's president of global affairs as saying, "with . . . hindsight, we feel that we overdid it a bit").

misinformation is not as politically palatable as it was only a few years ago. This did not happen by accident. The politicization of platforms' decisions has been a concerted project—one that has effectively built on arguments about the repressiveness of the public-health response to the pandemic and fears that social media platforms are biased against conservative viewpoints.<sup>16</sup>

But it would be a mistake to dismiss all this rhetoric as just politics. Criticism of platforms' moderation of medical misinformation has resonated because it reflects deeper and more universal concerns about the power and responsibility of social media platforms in the modern information economy. In drawing lines between what content is or is not allowed on their sites, platforms wield an enormous amount of power over public debate. This raises questions about how and why the exercise of such power is legitimate or in the public interest—debates that have been raging for years now.<sup>17</sup> It initially seemed like the pandemic made these questions easy to answer, at least with respect to medical misinformation: the public-health emergency justified platforms intervening to protect people from physical harm.<sup>18</sup> But questions about the legitimacy of platform power over public debate are never easy to answer, it turns out—even, or perhaps especially, in a public-health emergency. And as platforms' definition of the kinds of health-related claims they were willing to police expanded over the course of the pandemic to include claims that might be more properly understood as political claims related to health, or that had only an attenuated relationship to direct physical harm, platforms had a harder time defending their decisions. Bowing to political and public pressure (this time, from the opposite side of the political spectrum), the catch-all phrase “medical misinformation” became a vehicle for amorphous anxieties about any false speech related to COVID-19. Calls for platforms to do more often skipped over the important questions about whether platforms had the legitimacy to make the kinds of interventions being demanded.

These questions persist and still demand an answer. In what follows, I argue that instead of platforms' approach to content moderation blowing in the

---

16. Will Oremus, *How Social Media “Censorship” Became a Front Line in the Culture War*, WASH. POST (Oct. 9, 2022), <https://www.washingtonpost.com/technology/2022/10/09/social-media-content-moderation> [https://perma.cc/5KQS-TYWW].

17. John Bowers & Jonathan Zittrain, *Answering Impossible Questions: Content Governance in an Age of Disinformation*, 1 HARV. KENNEDY SCH. MISINFORMATION REV. 1, 5 (2020) (arguing that platforms need to “develop and build legitimacy around new ways of working through ambiguous and controversial content governance questions”); Evelyn Douek, *Governing Online Speech: From “Posts-As-Trumps” to Proportionality and Probability*, 121 COLUM. L. REV. 759, 820 (2021) (discussing platforms' need for legitimacy of their decision-making).

18. Evelyn Douek, *The Internet's Titans Make a Power Grab*, ATLANTIC (Apr. 18, 2020), <https://www.theatlantic.com/ideas/archive/2020/04/pandemic-facebook-and-twitter-grab-more-power/610213> [https://perma.cc/85YK-5P7Q].

political winds between overzealousness and repudiation, we should articulate an affirmative vision of the role that platforms can and should (and cannot and should not) play as gatekeepers of medical speech. In articulating that vision, First Amendment cases have something to teach us. Courts, after all, have also grappled with the difficult question of how to balance free-speech values with the value of facilitating access to knowledge. As this body of law suggests, falsity alone cannot and should not be enough to justify suppressing speech in public discourse – even in the context of health claims. Instead, intervention can be justified only where there is a clear relationship to specific and concrete harm or, as in the case of the medical profession’s self-regulation, in the context of particular relationships of vulnerability. Understanding how First Amendment law defines these relationships helps illuminate why overly aggressive moderation by platforms – which occupy a very different sociological role than medical professionals – not only will be ineffective but may even be counterproductive (as current political events suggest).

Indeed, widespread skepticism about platforms’ choices seems here to stay, and it will necessarily impact how responsible platforms should moderate content moving forward. The project of creating a healthy speech environment is not simply a question of determining ideal speech rules in a vacuum. As First Amendment doctrine recognizes, context matters, and the appropriate approach to speech regulation depends on particular sociological facts and relationships. The political environment – especially the institutional legitimacy and perceived trustworthiness of the decision maker – matters enormously to whether speech rules are accepted and effective. The problem is therefore cyclical: platforms overstepping their role during the pandemic opened them up to the critique that they were intervening in politics, a narrative that was then exploited for partisan gain, which in turn undermined users’ trust in platforms’ moderation decisions. That is, the politicization and delegitimization of platforms as trustworthy gatekeepers of medical truth not only makes them less enthusiastic about taking up that role, but also less effective at doing so.

The politicization of medical truth is a problem not only for platforms but for public health more broadly. Public trust and confidence in scientists declined sharply during the COVID-19 pandemic.<sup>19</sup> Platforms cannot simply content-moderate this problem away. Instead, we should be much more specific and cautious about the role we ask platforms to play in policing public debate. Debates about platforms’ moderation of medical misinformation are currently largely

---

19. Alec Tyson & Brian Kennedy, *Public Trust in Scientists and Views on Their Role in Policymaking*, PEW RSCH. CTR. (Nov. 14, 2024), <https://www.pewresearch.org/science/2024/11/14/public-trust-in-scientists-and-views-on-their-role-in-policymaking> [https://perma.cc/E7UB-LHVM].

polarized around two diametrically opposed views: they need to do much more, or they should not do anything at all. This Essay charts a different path forward. Part I tells a brief history of platforms' content moderation during the pandemic, their struggle to define the "medical misinformation" that they should suppress, and the political consequences that followed. Part II turns to First Amendment doctrine—not because platforms are required to follow it, but because it holds important lessons for the struggle of how to regulate medical truth and falsity. Part III then describes what these lessons mean for platforms—and for all of us, in our expectations of what content moderation can and should do. Medical misinformation is a pressing public-health challenge, but it is also now a political one. When trust in institutions is low, speech suppression is more likely to breed suspicion than inspire confidence. There is no content-moderation shortcut to the hard work of public education and trust building.

## I. CONTENT MODERATION AND THE COVID-19 PANDEMIC

As the Supreme Court confirmed just last Term in *Moody v. NetChoice*,<sup>20</sup> the First Amendment largely protects platforms' decisions about what speech to allow on their services.<sup>21</sup> That is, platforms have enormous discretion to pick whatever content-moderation rules they want. In their early years, while most platforms prohibited certain categories of content to make their services more pleasant and palatable for their users (and advertisers), they generally refused to take down content simply because it was false, even when it came to content about medical topics.<sup>22</sup> Sustained political pressure and public criticism about antivaccination content did eventually lead some platforms to take steps to reduce the circulation of such content on their services, but they mostly stopped short of outright removal of such claims.<sup>23</sup> Social media companies should not, platform executives insisted, be "arbiters of truth."<sup>24</sup>

---

20. 144 S. Ct. 2383 (2024).

21. *Id.* at 2401.

22. Douek, *supra* note 17, at 773-74.

23. Ryan Broderick, *Facebook Has Announced a Multistep Plan to Crack Down on Anti-Vax Misinformation*, BUZZFEED NEWS (Mar. 7, 2019, 2:00 PM), <https://www.buzzfeednews.com/article/ryanhatesthis/facebook-blocking-anti-vax-ads-reducing-newsfeed> [<https://perma.cc/S4Z4-SQ32>]; Christina Caron, *Pinterest Restricts Vaccine Search Results to Curb Spread of Misinformation*, N.Y. TIMES (Feb. 23, 2019), <https://www.nytimes.com/2019/02/23/health/pinterest-vaccination-searches.html> [<https://perma.cc/4NY9-WGYQ>].

24. Callum Borchers, *Twitter Executive on Fake News: "We Are Not the Arbiters of Truth,"* WASH. POST (Feb. 8, 2018, 3:20 PM EST), <https://www.washingtonpost.com/news/the-fix/wp/2018/02/08/twitter-executive-on-fake-news-we-are-not-the-arbiters-of-truth/> [<https://perma.cc/7ZRC-LH7Z>]; Supraja Srinivasan, *Tracking Veracity of Content a Tough Task: YouTube Executive*, ECON. TIMES (Mar. 24, 2018, 10:02 AM IST), <https://economictimes>

The COVID-19 pandemic, and fears about platforms' role in spreading the harmful misinformation that accompanied it, brought about dramatic reversals in policy, basically overnight.<sup>25</sup> As authorities warned of an “infodemic” of misinformation, social media platforms positioned themselves as part of the front-line response to harmful false claims.<sup>26</sup> Suddenly, many platforms became willing to remove certain COVID-19 related content that they judged to be false, breaking with their previous approach to policing medical misinformation and false claims more generally.

But these platforms made clear that their willingness to take down false speech about COVID-19 did not signal a more general willingness to start removing “misinformation” writ large. Instead, this newfound appetite to arbitrate the truth was limited to health-related claims in the context of a public-health emergency. Such content was, platforms insisted, simply different, for two reasons: first, false claims about COVID-19 were more likely to lead directly to physical harm;<sup>27</sup> and, second, the truth or falsity of such claims was more susceptible to verification by widely accepted, “authoritative” sources of information, such as the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC).<sup>28</sup> In other words, platforms justified

---

.indiatimes.com/industry/media/entertainment/media/tracking-veracity-of-content-a-tough-task-youtube-executive/articleshow/63439397.cms [https://perma.cc/AX3V-U5CG]; Mark Zuckerberg, FACEBOOK (Nov. 18, 2016), <https://www.facebook.com/zuck/posts/10103269806149061> [https://perma.cc/CQG4-BUX8].

25. Evelyn Douek, *The Internet's Titans Make a Power Grab*, ATLANTIC (Apr. 18, 2020), <https://www.theatlantic.com/ideas/archive/2020/04/pandemic-facebook-and-twitter-grab-more-power/610213> [https://perma.cc/PFN9-MPF4].
26. Julia Carrie Wong, *Tech Giants Struggle to Stem “Infodemic” of False Coronavirus Claims*, GUARDIAN (Apr. 10, 2020, 4:46 AM EDT), <http://www.theguardian.com/world/2020/apr/10/tech-giants-struggle-stem-infodemic-false-coronavirus-claims> [https://perma.cc/KRU4-FEFY].
27. Kang-Xing Jin, *Keeping People Safe and Informed About the Coronavirus*, META (Jan. 30, 2020), <https://about.fb.com/news/2020/12/coronavirus> [https://perma.cc/9374-QF4Q] (“We are doing this as an extension of our existing policies to remove content that could cause physical harm.”); Vijaya Gadde & Matt Derella, *An Update on Our Continuity Strategy During COVID-19*, X (formerly TWITTER) BLOG (Mar. 16, 2020), [https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html) [https://perma.cc/WUC4-LQA8] (“We’ll continue to prioritize removing content when it has a clear call to action that could directly pose a risk to people’s health or well-being . . .”).
28. See, e.g., Mark Zuckerberg, Remarks on Facebook Press Call 17 (Mar. 18, 2020), <https://about.fb.com/wp-content/uploads/2020/03/March-18-2020-Press-Call-Transcript.pdf> [https://perma.cc/8Y5E-TGHP] (“[T]he WHO for example . . . ha[s] broad trust and a government mandate on [COVID-19] in a way that in other domains there just (isn’t) something like that.”); *Medical Misinformation Policy*, YOUTUBE HELP, <https://support.google.com/youtube/answer/9891785> [https://perma.cc/CV77-7SX2] (“YouTube doesn’t allow content that poses a serious risk of egregious harm by spreading medical misinformation that

taking down health misinformation on the basis that it was not merely a matter of politics or opinion, but instead a matter of fact and expertise. These kinds of claims were, in the words of Mark Zuckerberg at the time, “just in a different class” than other kinds of misinformation, because of the “imminent risk of danger” and the fact that “it’s easier to set policies that are a little more black and white.”<sup>29</sup>

The praise for these moves was swift and widespread—a noticeable change in tune from the years of persistent public criticism of platforms that preceded it. “Facebook [i]s [m]ore [t]rustworthy [t]han the President,” one headline declared.<sup>30</sup> “Has the Coronavirus [k]illed the Techlash?” mused another.<sup>31</sup> But the honeymoon period did not last long. The neat distinction platforms drew between health misinformation and other kinds of false claims quickly broke down, for several reasons.

First, identifying good and bad information in the context of a public-health emergency caused by a novel virus was (predictably) more difficult than platforms acknowledged. As is to be expected in such circumstances, the guidance from public-health authorities was fluid, constantly changing, and sometimes contradictory.<sup>32</sup> To take one of the most infamous examples, during the early stages of the pandemic, institutions such as WHO and the U.S. Surgeon General told the public that masks were not necessary.<sup>33</sup> A few months later, mask mandates were widespread.<sup>34</sup> In trying to keep up with this shifting guidance, platforms that had initially banned ads for masks on their services had to reverse

---

contradicts local health authorities’ (LHAs) or the World Health Organization’s (WHO) guidance about specific health conditions and substances.”).

29. Ben Smith, *When Facebook Is More Trustworthy than the President*, N.Y. TIMES (Mar. 15, 2020), <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html> [https://perma.cc/TM69-3ERK].

30. *Id.*

31. Levy, *supra* note 13.

32. See Vivek H. Murthy, *Confronting Health Misinformation: The U.S. Surgeon General’s Advisory on Building a Healthy Information Environment*, DEP’T OF HEALTH & HUM. SERVS. 4 (2021), <https://www.hhs.gov/sites/default/files/surgeon-general-misinformation-advisory.pdf> [https://perma.cc/EH9D-ZWQW] (“[S]cientific knowledge about COVID-19 has evolved rapidly over the past year, sometimes leading to changes in public health recommendations.”).

33. Maria Cramer & Knvul Sheikh, *Surgeon General Urges the Public to Stop Buying Face Masks*, N.Y. TIMES (Sept. 12, 2021), <https://www.nytimes.com/2020/02/29/health/coronavirus-n95-face-masks.html> [https://perma.cc/CTT9-J268].

34. Sarah Mervosh, Manny Fernandez & Campbell Robertson, *Mask Rules Expand Across U.S. as Clashes over the Mandates Intensify*, N.Y. TIMES (Aug. 3, 2020), <https://www.nytimes.com/2020/07/16/us/coronavirus-masks.html> [https://perma.cc/KG37-GPZX].

their policies.<sup>35</sup> The earlier insistence that “it’s easier to set policies that are a little more black and white” became harder to maintain. Because platforms (like the public-health authorities themselves) did not effectively communicate their reasons for these reversals, such backtracking simply became proof to critics that attempts to stamp out false claims come with the inherent risk of stamping out valuable true claims.<sup>36</sup>

Second, these dynamics were both intentionally exploited and unintentionally exacerbated as claims about the virus itself became highly politicized, undermining any easy separation between the political and medical spheres. Most prominently, President Donald Trump consistently made false claims about the pandemic, including promoting ineffective (or dangerous) “miracle cures” and politicizing individual precautionary measures like mask wearing.<sup>37</sup> Expressing agreement or disagreement with such claims by political figures became as much a proxy for support for that candidate as a signal of belief or disbelief in the underlying medical claim.<sup>38</sup> As a result, platforms found themselves in the position of appearing to take sides in a political debate when they moderated these kinds of claims. Predictably, platforms attracted criticism both when they left political figures’ false claims up<sup>39</sup> and when they took them down.<sup>40</sup>

- 
35. See, e.g., Rob Leathern, *Allowing the Promotion of Non-Medical Masks on Facebook*, FACEBOOK BUS. (June 10, 2020), <https://www.facebook.com/business/news/allowing-the-promotion-of-non-medical-masks-on-facebook> [https://perma.cc/Q2LC-J88R].
  36. See, e.g., F.D. Flam, *Facebook, YouTube Erred in Censoring Covid-19 ‘Misinformation,’* BLOOMBERG (June 7, 2021, 9:00 AM EDT), <https://www.bloomberg.com/opinion/articles/2021-06-07/facebook-youtube-erred-in-censoring-covid-19-misinformation> [https://perma.cc/JX8L-YMDD].
  37. See Dylan Scott, *Trump Has Been the Biggest Source of Covid-19 Misinformation, Study Finds*, VOX (Oct. 5, 2020, 11:50 AM EDT), <https://www.vox.com/coronavirus-covid19/21497221/donald-trump-covid-19-coronavirus-news-misinformation-study> [https://perma.cc/6ULB-C4K7].
  38. See, e.g., Cailin O’Connor & James Owen Weatherall, *Hydroxychloroquine and the Political Polarization of Science*, BOS. REV. (May 4, 2020), <https://www.bostonreview.net/articles/cailin-oconnor-james-owen-weatherall-covid-19-and-polarization> [https://perma.cc/6UYH-S7GM].
  39. See, e.g., Cristiano Lima, *Democrats Renew Calls for Twitter to Suspend Trump for Spreading Misinformation*, POLITICO (Nov. 4, 2020, 2:54 PM EST), <https://www.politico.com/news/2020/11/04/democrats-twitter-trump-misinformation-434070> [https://perma.cc/3H4W-9U5A].
  40. See, e.g., Press Release, Off. of Sen. Ron Johnson, *Sen. Johnson Demands YouTube Explain COVID-19 Censorship Policies* (Sep. 22, 2022), <https://www.ronjohnson.senate.gov/2022/9/sen-johnson-demands-youtube-explain-covid-19-censorship-policies> [https://perma.cc/74R8-FZRA].

Third, the opaque relationships that major platforms had with government actors compounded fears of political bias. These platforms worked with federal officials at the White House, the Office of the Surgeon General, and the CDC to give and receive information about what people were saying on social media and to discuss how to combat misinformation.<sup>41</sup> But communication to the public about these relationships with government actors was muddy at best and obfuscatory at worst. Platforms did not have a clear message about how they were drawing a line between “deferring to authoritative sources” in determining the facts about COVID-19 and allowing the government to call the shots. This in turn exposed them to criticism that they were suppressing both valuable information and critiques of the government’s response to the pandemic, all at the behest of the officials they were working with.<sup>42</sup> Platforms insisted that their content-moderation decisions were always made independently,<sup>43</sup> but the lack of transparency into government-platform communications raised the specter that these relationships in fact allowed the government to make exactly the kind of speech-related decisions that the First Amendment places off limits.<sup>44</sup>

Fourth, despite early praise for platforms’ efforts, pressure for platforms to do more, more, *more* about the problem of medical misinformation remained unrelenting, leading platforms to expand what they were willing to take down. Platforms bowed to this pressure, no doubt in part because they did not want to be responsible for harm, but also likely in part to avoid political costs. Whatever the reason, many platforms started to expand the category of “medical misinformation” they were willing to police beyond its originally limited contours, undermining the reasons they had pointed to for treating COVID-19 misinformation differently than other kinds of misinformation. Platforms had initially justified their extraordinary moderation in this context on the basis of a direct

---

41. See *Murthy v. Missouri*, 144 S. Ct. 1972, 1982–83 (2024).

42. See Vivek Ramaswamy & Jed Rubenfeld, Opinion, *Twitter Becomes a Tool of Government Censorship*, (Aug. 17, 2022, 1:47 PM ET), <https://www.wsj.com/articles/twitter-becomes-a-tool-of-government-censors-alex-berenson-twitter-facebook-ban-covid-misinformation-first-amendment-psaki-murthy-section-230-antitrust-11660732095> [<https://perma.cc/VT7C-BZMF>].

43. See, e.g., Letter from Mark Zuckerberg, *supra* note 15 (“Ultimately, it was our decision whether or not to take content down, and we own our decisions . . .”); see also *Murthy*, 144 S. Ct. at 1987–88 (“[The] evidence indicates that the platforms had independent incentives to moderate content and often exercised their own judgment.”).

44. Yoel Roth, *Getting the Facts Straight: Some Observations on the Fifth Circuit Ruling in Missouri v. Biden*, KNIGHT FIRST AMEND. INST. AT COLUM. UNIV. (Sep. 27, 2023), <http://knightcolumbia.org/blog/getting-the-facts-straight-some-observations-on-the-fifth-circuit-ruling-in-missouri-v-biden-1> [<https://perma.cc/6KU4-NH6N>] (pointing to the “First Amendment concerns” related to the “jawboning of social media companies” by government actors and calling for “radical transparency” to combat distrust).

link between COVID-19 misinformation and physical harm. They therefore focused on removing medical claims that led people to act in ways that put them at serious direct risk of physical harm – such as advocating ingesting a harmful “cure” or exposing themselves to the virus – in situations where the recipient did not have the capacity to properly evaluate that risk. But politicians did not have such a limited definition of the kinds of content platforms should remove.<sup>45</sup> As a result, the vague term “medical misinformation” came to encompass a far broader range of false claims about COVID-19, including unverified claims about the origins and nature of the virus and its spread that had a much more attenuated relationship to physical harm. For example, some platforms removed claims that 5G technology networks were responsible for the virus’s rapid spread.<sup>46</sup> Others removed posts that suggested that the virus was manufactured.<sup>47</sup> Beliefs in such false claims may well lead people to act in misguided ways – vandalizing cell towers that they believed transmitted harmful 5G radio waves, for example<sup>48</sup> – but the same might be said to be true of many forms of false (or, indeed, true) speech. A person that sets fire to a cell tower knows that they are causing property damage. But someone who ingests a substance under the assumption that it will provide a cure but instead suffers physical harm is very differently situated. The direct link to physical harm that platforms had pointed to as justifying their willingness to moderate COVID-19 misinformation simply did not apply to the expanded category of claims that platforms started to remove.

- 
45. See Isabella Grullón Paz, *Elizabeth Warren Asks Amazon to ‘Stop Peddling Misinformation About Covid Vaccines and Treatments.’*, N.Y. TIMES (Sept. 24, 2021), <https://www.nytimes.com/2021/09/08/world/elizabeth-warren-amazon-covid-misinformation.html> [<https://perma.cc/KNW7-3V5B>]; Monika Bickert, *How We’re Taking Action Against Vaccine Misinformation Superspreaders*, META (Aug. 18, 2021), <https://about.fb.com/news/2021/08/taking-action-against-vaccine-misinformation-superspreaders> [<https://perma.cc/3DBU-KT68>] (pushing back on calls to ban certain accounts that “are not posting content that breaks our rules”).
46. Adam Satariano & Davey Alba, *Burning Cell Towers, Out of Baseless Fear They Spread the Virus*, N.Y. TIMES (Apr. 11, 2020), <https://www.nytimes.com/2020/04/10/technology/coronavirus-5g-uk.html> [<https://perma.cc/BMC4-VRGX>] (describing the removal of 5G conspiracy posts by social-media platforms); Rebecca Heilweil, *How the 5G Coronavirus Conspiracy Theory Went from Fringe to Mainstream*, VOX (Apr. 24, 2020, 6:20 AM CDT), <https://www.vox.com/recode/2020/4/24/21231085/coronavirus-5g-conspiracy-theory-covid-facebook-youtube> [<https://perma.cc/UE3T-UXCT>] (describing Facebook’s removal of anti-5G groups and content).
47. Alex Hern, *Facebook Lifts Ban on Posts Claiming Covid-19 Was Man-Made*, GUARDIAN (May 27, 2021, 10:58 AM EDT), <http://www.theguardian.com/technology/2021/may/27/facebook-lifts-ban-on-posts-claiming-covid-19-was-man-made> [<https://perma.cc/HD4H-WEBM>] (describing Facebook’s four-month ban on the claim that COVID-19 was man-made, which exposed some users to the risk of having their accounts banned entirely).
48. Satariano & Alba, *supra* note 46.

To make matters worse, platforms defined the category of COVID-19 misinformation somewhat differently from one another. Speech that was not allowed on some platforms (the lab-leak theory of the virus' origin on Facebook, for example) might be permitted on another (like Twitter or YouTube).<sup>49</sup> Rather than alleviating concerns about platform censorship, however, this only intensified them. On the one hand, public discussion about content moderation often did not distinguish between platforms' approaches. The fact that Meta removed the lab-leak theory while Twitter did not, for example, did not prevent people from using it as an example of platforms' overreach in general.<sup>50</sup> On the other hand, the fact that platforms came to different conclusions about the best approach underlined the subjective nature of the judgments that they had insisted were made on the basis of authoritative guidance. As a result, the mission creep in the expanding definition of "medical misinformation" gave further oxygen to criticisms that platforms were not just removing claims on the basis that they were harmful to people's health, but going further to please government critics or to shore up the legitimacy and authority of the public-health institutions they were working with.<sup>51</sup>

Debates about content moderation flattened other important nuances, too. Content moderation at the scale of online social media platforms always involves mistakes, including both false positives and false negatives – and platforms' enforcement of their medical misinformation policies was no different.<sup>52</sup> Indeed, platforms had warned that they would make more enforcement errors during the pandemic because they were forced to rely more heavily on automated moderation with reduced human oversight.<sup>53</sup> Such mistakes, however, were easy to exploit for those seeking proof of platforms' intentional censorship. Similarly, platforms tried to strike a balance between speech and safety by using measures

---

49. See, e.g., Cristiano Lima, *Facebook No Longer Treating "Man-Made" Covid as a Crackpot Idea*, POLITICO (May 26, 2021), <https://www.politico.com/news/2021/05/26/facebook-ban-covid-man-made-491053> [<https://perma.cc/26YP-C694>] (describing platforms' different approaches to moderating the COVID-19 lab-leak theory).

50. Jonathan Turley, *COVID Lab Leak Is a Scandal of Media and Government Censorship*, N.Y. POST (Feb. 27, 2023, 9:21 AM ET), <https://nypost.com/2023/02/26/covid-lab-leak-is-a-scandal-of-media-and-government-censorship> [<https://perma.cc/WK2M-YELA>] (discussing Twitter's content moderation in the context of the lab-leak theory as an example of over-zealous content moderation).

51. See Editorial Board, *Facebook's Lab-Leak About-Face*, WALL ST. J. (May 27, 2021, 6:23 PM EDT), <https://www.wsj.com/articles/facebooks-lab-leak-about-face-11622154198> [<https://perma.cc/C2GE-YAWC>].

52. Taylor Lorenz, *Twitter Labeled Factual Information About Covid-19 as Misinformation*, WASH. POST (Aug. 25, 2022), <https://www.washingtonpost.com/technology/2022/08/25/twitter-factual-covid-info-labeled-misinformation> [<https://perma.cc/79HV-PAWT>].

53. Douek, *supra* note 17, at 802.

short of taking content down, such as adding fact-checking labels or preventing certain posts from being algorithmically amplified.<sup>54</sup> Such measures certainly have an impact on the reach of content, but do not suppress speech in the same way as outright removals of posts. Nonetheless, these measures, too, were decried as censorship.

All of these dynamics meant platforms' content-moderation practices became a central battleground in the emerging culture wars.<sup>55</sup> Politicians on the left continued to lament the inadequacy of platforms' efforts—a frustrated President Joe Biden accused social media platforms of “killing people” by allowing vaccine misinformation to spread.<sup>56</sup> He later walked back this claim,<sup>57</sup> but the fact that he made it in the first place only demonstrates how inflammatory political rhetoric about platforms' level of responsibility had become by that point in the pandemic. Meanwhile, conservative politicians thought platforms were doing far too much—they decried social media “censorship” and expanded on years-long allegations (repeatedly debunked<sup>58</sup>) that platforms were biased against conservative viewpoints.<sup>59</sup>

So today, a mere few years after platforms first embraced their role as curators of medical truth, platforms' belief that there is a category of false claims called “medical misinformation” that they can safely moderate without becoming embroiled in politics looks far too optimistic. What the past few years have

---

54. See, e.g., Yoel Roth & Nick Pickles, *Updating Our Approach to Misleading Information*, TWITTER BLOG (May 11, 2020), [https://blog.x.com/en\\_us/topics/product/2020/updating-our-approach-to-misleading-information](https://blog.x.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information) [<https://perma.cc/MQB4-9S42>].

55. Oremus, *supra* note 16.

56. Zolan Kanno-Youngs & Cecilia Kang, *‘They’re Killing People’: Biden Denounces Social Media for Virus Disinformation*, N.Y. TIMES (July 19, 2021), <https://www.nytimes.com/2021/07/16/us/politics/biden-facebook-social-media-covid.html> [<https://perma.cc/BZF3-B4DW>].

57. Eugene Scott & Rachel Lerman, *Biden Clarifies Comments About Facebook ‘Killing People’ with Vaccine Misinformation*, WASH. POST (July 19, 2021, 2:33 PM EDT), <https://www.washingtonpost.com/politics/2021/07/19/biden-facebook-misinformation> [<https://perma.cc/9ZQY-UNYN>].

58. See, e.g., Mohsen Mosleh et al., *Differences in Misinformation Sharing Can Lead to Politically Asymmetric Sanctions*, 634 NATURE 609, 609 (2024); Paul M. Barrett & J. Grant Sims, *False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives*, NYU STERN CTR. FOR BUS. & HUM. RTS. 1 (Feb. 2021), [https://bhr.stern.nyu.edu/wp-content/uploads/2024/02/NYUFalseAccusation\\_2.pdf](https://bhr.stern.nyu.edu/wp-content/uploads/2024/02/NYUFalseAccusation_2.pdf) [<https://perma.cc/ZM2W-MWBA>].

59. See, e.g., Jessi Turnure, *Senate Republicans Slam Big Tech over ‘Censorship’ of COVID-19 Origins*, NEWS10 (June 10, 2021, 5:44 PM EDT), <https://www.news10.com/washington/washington-dc/senate-republicans-slam-big-tech-over-censorship-of-covid-19-origins> [<https://perma.cc/PDY9-VVKA>]; Alayna Treene, *Over 60 Conservatives Demand Big Tech End ‘Censorship’*, AXIOS (May 1, 2018), <https://www.axios.com/2018/05/01/over-60-prominent-1525192042> [<https://perma.cc/Q8G8-CZ64>].

made clear is that, one way or another, all misinformation is or can become political misinformation, too.

It is possible that there was no good answer for platforms caught in this bind. With the benefit of hindsight, and the knowledge of all the political weaponization of platforms' actions that followed, perhaps it is too easy to argue that platforms erred. At the time, public discourse was dominated by fears that platforms' failure to remove false claims led to significant excess deaths. "Free speech" concerns might have seemed overly abstract in the context of such a pressing public-health emergency.<sup>60</sup> But this is precisely why it is important to reflect now on the missteps that were made — because calls for speech suppression are often loudest and hardest to resist in moments of crisis. Understanding why and how that happened is the only way to learn from the experience for the future.

In distilling lessons from this experience, we can turn to existing frameworks. Platforms are not the first speech regulators to wrestle with the competing equities involved in the dissemination of medical knowledge. First Amendment doctrine has also grappled with this tension — between acknowledging that there is such a thing as medical expertise that people need to be able to rely upon, on the one hand, and recognizing the dangers of giving any authority the power to punish dissent, on the other. What the First Amendment cases teach us is that "medical misinformation" is not a special category to which the ordinary reasons for caution about speech suppression do not apply. While there are certain circumstances in which false medical claims can be punished, these are defined by very particular harms or vulnerabilities, rather than by the mere fact of falsity. The next Part explores those lessons.

## II. MEDICAL MISINFORMATION AND THE FIRST AMENDMENT

There are all sorts of reasons why platforms will not and should not adopt First Amendment standards in writing their content-moderation rules, not least

---

60. See, e.g., Felicia Sonmez & Cat Zakrzewski, *Surgeon General Vivek Murthy Issues Advisory on Dangers of Health Misinformation amid Rise in Coronavirus Cases*, WASH. POST (July 15, 2021), [https://www.washingtonpost.com/politics/surgeon-general-vivek-murthy-issues-advisory-on-dangers-of-health-misinformation-amid-rise-in-covid-19-cases/2021/07/15/a3e6d5ae-e591-11eb-8aa5-5662858b696e\\_story.html](https://www.washingtonpost.com/politics/surgeon-general-vivek-murthy-issues-advisory-on-dangers-of-health-misinformation-amid-rise-in-covid-19-cases/2021/07/15/a3e6d5ae-e591-11eb-8aa5-5662858b696e_story.html) [<https://perma.cc/L3KB-FU9W>]; Press Release, Off. of Rep. Lori Trahan, Trahan, Schiff Urge Meta to Continue Removing Dangerous COVID-19 Misinformation (Aug. 3, 2022), <https://trahan.house.gov/news/document-single.aspx?DocumentID=2578> [<https://perma.cc/77C5-BN7Q>]; Press Release, Off. of Sen. Mark R. Warner, Warner, Menendez, Hirono Slam Facebook for Spread of Vaccine Misinformation (June 22, 2021), <https://www.warner.senate.gov/public/index.cfm/2021/6/warner-menendez-hirono-slam-facebook-for-spread-of-vaccine-misinformation> [<https://perma.cc/89AW-6V2Y>].

that user and advertiser preferences mean that such rules will be very bad for business.<sup>61</sup> Nevertheless, the core principles underpinning First Amendment doctrine hold important insights about the difficult project of speech regulation more generally—insights that also can be relevant to the private systems of speech regulation that platforms create when they engage in content moderation. And indeed, in writing and applying their content-moderation policies, platforms have been heavily influenced by the First Amendment tradition.<sup>62</sup>

One of the most obvious ways in which platforms have been influenced by First Amendment law is in their early reticence to remove false claims.<sup>63</sup> It is a central principle of First Amendment law that, generally speaking, the government cannot punish people for saying things that are wrong. In one of the most famous sentences in the First Amendment canon, Justice Holmes declared that “the best test of truth is the power of the thought to get itself accepted in the competition of the market.”<sup>64</sup> This sentence, and the marketplace-of-ideas metaphor that it gave rise to, embodies the idea that law should not seek to fix the line between truth and falsity, and that collective knowledge is instead better advanced through the rough and tumble of public discourse and debate. Therefore, in the words of Justice Brandeis, the remedy to false speech “is more speech, not enforced silence.”<sup>65</sup>

This theory has profoundly influenced First Amendment doctrine.<sup>66</sup> The Supreme Court echoed Justices Holmes and Brandeis nearly a century later, in *United States v. Alvarez*, when it rejected the argument that false speech is presumptively unprotected by the First Amendment.<sup>67</sup> In striking down a law that made it a crime to lie about receiving the Congressional Medal of Honor, Justice Kennedy’s opinion for the Court argued that upholding such a law would “endorse government authority to compile a list of subjects about which false statements are punishable.”<sup>68</sup> Invoking George Orwell’s dystopian novel *1984*, Justice Kennedy concluded that such a power is inconsistent with a free society and that

---

61. Cf. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1625–30 (2018) (summarizing platforms’ reasons for engaging in content moderation, despite the fact that they have no legal obligation to do so).

62. *Id.* at 1621.

63. Douek, *supra* note 17, at 773–74.

64. *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

65. *Whitney v. California*, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring).

66. Joseph Blocher, *Free Speech and Justified True Belief*, 133 HARV. L. REV. 439, 440–41 (2019).

67. 567 U.S. 709, 727 (2012) (“The remedy for speech that is false is speech that is true. This is the ordinary course in a free society.”).

68. *Id.* at 723.

“[o]ur constitutional tradition stands against the idea that we need Oceania’s Ministry of Truth.”<sup>69</sup>

*Alvarez* is the Court’s most recent and most important opinion on the regulation of lies, and it makes clear that even false speech is protected by the First Amendment. As a general proposition, this conclusion is hard to object to—a right to free speech would mean little if the government could silence people simply by declaring that they said something untrue. After all, the point of guaranteeing a right to free speech is to allow citizens to say things even when the government does not want them to. But this general principle only goes so far, because all systems of free expression also recognize that the principle of free speech has limits. Certainly, when it comes to false speech, it would be a caricature to suggest that the First Amendment prohibits *all* government punishment of lies and misinformation.<sup>70</sup> As *Alvarez* itself recognizes, the government punishes lies all the time. In his majority opinion, Justice Kennedy acknowledged—and disclaimed the constitutional vulnerability of—laws that prohibit false statements to government officials, the impersonation of government officials, perjury, “defamation, fraud, or [cases involving] some other legally cognizable harm associated with a false statement.”<sup>71</sup>

The lesson of *Alvarez*, then, is one that platforms echoed in their statements early in the pandemic: lies may be punishable in certain circumstances, but there are good reasons to define those circumstances narrowly. Platforms, as nonstate actors, need not be constrained by the especially high bar that the First Amendment demands regulators meet before lies can be punishable.<sup>72</sup> But what *Alvarez* rightly makes clear is that there are no subject categories (like “medical misinformation”) that are per se exempt from the general disciplining requirement to show that punishable lies must be narrowly defined and directly linked to cognizable harm. This is because, whatever the subject matter, prohibitions on false speech are vulnerable to being used to advance political aims or suppress criticism.

---

69. *Id.* (citing GEORGE ORWELL, NINETEEN EIGHTY-FOUR (1949)).

70. Here I gratefully draw on ideas and arguments made with Genevieve Lakier in Evelyn Douek & Genevieve Lakier, *Rereading Alvarez*, KNIGHT FIRST AMEND. INST. AT COLUMBIA UNIV. (May 18, 2022), <http://knightcolumbia.org/blog/rereading-alvarez> [<https://perma.cc/22ZR-7K4Q>] (citing examples of government prohibitions against lying when speaking to police or applying for a government benefit or status).

71. *Alvarez*, 569 U.S. at 720–21.

72. See, for example, the (in)famous case *Winter v. G.P. Putnam’s Sons*, 938 F.2d 1033, 1036 (9th Cir. 1991), which held that a book publisher could not be sued by plaintiffs who became severely ill after relying on an (apparently misleading) reference guide on mushrooms because, in part, the First Amendment required that publishers should not be liable for “the ideas and expression contained in a book.”

The Supreme Court’s decision last Term in *Murthy v. Missouri* is a clear illustration of courts’ reluctance to apply different First Amendment rules to the regulation of false speech depending on its falsity.<sup>73</sup> The case involved sprawling allegations that during the 2020 election season and the COVID-19 pandemic, members of the Biden Administration had unconstitutionally pressured (or “jawboned”<sup>74</sup>) social media platforms into removing what the government officials thought was election and medical misinformation from their services.<sup>75</sup> The Supreme Court ultimately dismissed the case for lack of standing,<sup>76</sup> but only after courts at every prior stage of the case—from the lower courts to oral argument at the Supreme Court—had implicitly rejected the idea that the particular expertise of the government actor, or the kind of speech they were targeting, was relevant to the legal analysis.<sup>77</sup> Instead, they invoked the same doctrinal principles, regardless of whether they were discussing government pressure to remove election-related misinformation and foreign-influence campaigns or COVID-19 misinformation. No one suggested, for example, that the CDC deserved greater deference or latitude in its communications with platforms than the FBI because of the CDC’s medical expertise or the health-related (rather than election-related) nature of their remit.

The refusal to treat different kinds of false speech differently, depending on their subject matter, is a very good thing, insofar as it means that government actors cannot aggrandize power simply by reframing the topic of the speech they seek to regulate. A recent example from Florida provides a stark illustration of how such power could be abused. The Floridian Department of Health wrote to television stations demanding they not run political ads in support of a constitutional amendment that would protect abortion access in the state. It argued that the ads were a “sanitary nuisance” because they contained false claims and “would likely have a detrimental effect on the lives and health of pregnant

---

73. 144 S. Ct. 1972 (2024).

74. See, e.g., Genevieve Lakier, *Informal Government Coercion and the Problem of “Jawboning,”* LAWFARE (July 26, 2021, 3:52 PM), <https://www.lawfareblog.com/informal-government-coercion-and-problem-jawboning> [https://perma.cc/BE2X-XMQZ] (defining “jawboning” as the government’s use of “informal means, rather than democratically enacted laws” to encourage others to suppress speech).

75. *Murthy*, 144 S. Ct. at 1981.

76. *Id.*

77. See, e.g., *Missouri v. Biden*, 83 F.4th 350, 392 (5th Cir. 2023) (analyzing the actions of the White House, the Surgeon General, the Centers for Disease Control, the FBI, and the Cybersecurity and Infrastructure Security Agency as subject to the same First Amendment constraints); Transcript of Oral Argument at 54–56, *Murthy*, 144 S. Ct. 1972 (No. 23–411) (discussing the wide variety of topics on which government actors communicate with platforms).

women in Florida.”<sup>78</sup> As the district court acknowledged in concluding that the Department violated the First Amendment rights of the organization that wanted to run the ads, “if the State can re-brand rank viewpoint discriminatory suppression of political speech as a ‘sanitary nuisance,’ then any political viewpoint with which the State disagrees is fair game for censorship.”<sup>79</sup> As this suggests, speech about medical topics and political speech are not clearly distinct categories. For this reason, claims of medical misinformation can also be a powerful tool of government censorship. And indeed, during the pandemic, governments around the world used public-health concerns to justify imposing speech laws that suppressed criticisms of those governments.<sup>80</sup>

That said, the First Amendment does recognize that medical speech is not always like other speech.<sup>81</sup> While health-related speech is treated like any other speech when uttered in public discourse, there are special contexts in which the First Amendment does permit greater regulation. But rather than justifying further intervention by platforms, the rationales for these special rules underline the reasons why platforms have struggled to gain legitimacy as regulators of medical claims.

First, false or misleading health claims made in the context of speech attempting to sell a particular product receive no constitutional protection.<sup>82</sup> This kind of deception is regulated by, for example, the Food and Drug Administration and the Federal Trade Commission. However, punishing lies in these contexts is just a subset of the First Amendment’s commercial-speech doctrine, rather than a recognition of any particular characteristics of health-related speech. That doctrine rests on the particular reliance that consumers have on sellers to speak honestly about their products and the knowledge asymmetry inherent in that relationship.<sup>83</sup>

---

78. *Floridians Protecting Freedom, Inc. v. Ladapo*, No. 24CV419, 2024 WL 4518291, at \*1 (N.D. Fla. Oct. 17, 2024).

79. *Id.* at \*5.

80. *Covid-19 Triggers Wave of Free Speech Abuse*, HUM. RTS. WATCH (Feb. 11, 2021, 3:00 AM EST), <https://www.hrw.org/news/2021/02/11/covid-19-triggers-wave-free-speech-abuse> [<https://perma.cc/5V9W-MZ9D>]; *Covid-19: Global Attack on Freedom of Expression Is Having a Dangerous Impact on Public Health Crisis*, AMNESTY INT’L (Oct. 19, 2021), <https://www.amnesty.org/en/latest/news/2021/10/covid-19-global-attack-on-freedom-of-expression-is-having-a-dangerous-impact-on-public-health-crisis> [<https://perma.cc/3FEU-NJGW>].

81. See Claudia E. Haupt, *The Limits of Professional Speech*, YALE L.J.F. 185, 188 (2018) (“[Professional speech] is treated differently under the First Amendment than other types of speech . . .”).

82. See Jane R. Bambauer, *Snake Oil Speech*, 93 WASH. L. REV. 73, 101-02 (2018).

83. See *Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc.*, 425 U.S. 748, 772 n.24 (1976) (“The truth of commercial speech, for example, may be more easily verifiable by its disseminator than, let us say, news reporting or political commentary, in that ordinarily the

Second, and more significantly, doctors can be subject to regulation based on what they say in certain contexts, even though such regulation would be presumptively unconstitutional if applied to other speakers. Medical-licensing laws act as a prior restraint on doctors' speech by allowing them to provide medical advice to patients only after they have received a license from the state, and doctors that give bad advice can be sanctioned on the basis of the content of what they say.<sup>84</sup> Indeed, "[w]ithout so much as a nod to the First Amendment, doctors are routinely held liable for malpractice" because they say the wrong thing, or fail to say the right thing.<sup>85</sup>

The fact that doctors can be punished for providing incorrect information to patients is in obvious tension with the rest of First Amendment doctrine that generally prohibits punishing people for saying things that are wrong or false. For this reason, the limits of this kind of quasi-governmental regulation of falsity are narrowly defined.<sup>86</sup> While doctors can be sanctioned for what they say "in the course of professional practice" — which is to say, in the context of a doctor-patient relationship<sup>87</sup> — they retain their First Amendment rights as citizens when engaging in public debate and when talking to the public at large.<sup>88</sup> This results in what Claudia Haupt has aptly named the "Dr. Oz paradox" — the fact that "the law sanction[s] giving bad advice to *one* patient, while it permits giving

---

advertiser seeks to disseminate information about a specific product or service that he himself provides and presumably knows more about than anyone else.").

84. See Haupt, *supra* note 81, at 188.

85. Robert Post, *Informed Consent to Abortion: A First Amendment Analysis of Compelled Physician Speech*, 2007 U. ILL. L. REV. 939, 950.

86. See Haupt, *supra* note 81, at 188.

87. Post, *supra* note 85, at 947 (quoting Daniel Halberstam, *Commercial Speech, Professional Speech, and the Constitutional Status of Social Institutions*, 147 U. PA. L. REV. 771, 843 (1999)); Claudia E. Haupt, *Professional Speech*, 125 YALE L.J. 1238, 1254-56 (2016).

88. See *Lowe v. SEC*, 472 U.S. 181, 232 (1985) (White, J., concurring) ("Where the personal nexus between professional and client does not exist, and a speaker does not purport to be exercising judgment on behalf of any particular individual with whose circumstances he is directly acquainted, government regulation ceases to function as legitimate regulation of professional practice with only incidental impact on speech; it becomes regulation of speaking or publishing as such, subject to the First Amendment's command that 'Congress shall make no law . . . abridging the freedom of speech, or of the press . . .'" (quoting U.S. CONST. amend. I)).

bad advice to *millions* of YouTube or television viewers.”<sup>89</sup> (Coincidentally, Dr. Oz has also been tapped to join the next Trump administration.<sup>90</sup>)

This is a paradox because it is so counterintuitive. We might think that the law should recognize the greater capacity for harm arising from false advice that is widely distributed rather than given to a single person. But this gets it backwards. First Amendment doctrine has been reluctant to hinge the constitutionality of speech regulation in the public sphere writ large on governmental or judicial assessment that speech may be generally bad or harmful if people are persuaded to act upon it in the future,<sup>91</sup> in no small part because of the historical track record of the government getting that assessment startlingly and dangerously wrong.<sup>92</sup> Thus, it is the regulation of the speech of medical professionals directed toward individual patients that is the exception to the general rule, and only because it happens outside of the bounds of regular public debate. This disparate treatment of doctors’ speech is doctrinally justified by the unique characteristics and context of the communicative act between a doctor and their patient: the particular relationship of vulnerability that a patient has with a doctor; the knowledge asymmetry between them; the doctor’s obligation not to advance their own interests but to act in the best interests of the patient; and the doctor’s responsibility as a representative and conduit of the insights of the broader medical community.<sup>93</sup>

The last point is crucial—the content-based standards that demarcate the line for liability are generally not fixed by the government, but by the profession itself.<sup>94</sup> In medical-malpractice suits, for example, a doctor is measured against the norms established by the medical profession, rather than some externally imposed standard.<sup>95</sup> As Haupt has explained, it is this existence of a learned profession—a “knowledge community” on whose behalf the individual doctor

---

89. Claudia E. Haupt, *Pseudoprofessional Advice*, 103 B.U. L. REV. 775, 783 (2023) (quoting Claudia E. Haupt, *The Dr. Oz Paradox*, HARV. L. PETRIE-FLOM CTR.: BILL OF HEALTH (Sept. 21, 2022), <https://blog.petrieflom.law.harvard.edu/2022/09/21/the-dr-oz-paradox> [<https://perma.cc/G2MG-ZA7Z>]).

90. Weiland et. al, *supra* note 12.

91. See, e.g., *United States v. Stevens*, 559 U.S. 460, 470 (2010) (noting that “[t]he First Amendment’s guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits” and declaring such a suggestion otherwise “startling and dangerous”).

92. See Christina E. Wells, *Fear and Loathing in Constitutional Decision-Making*, 115 WIS. L. REV. 117, 117 (2005) (“Time and again, courts, including the Supreme Court, have deferred to questionable, if not outright illegitimate, government actions . . .”).

93. Haupt, *supra* note 87, at 1253–54.

94. *Id.* at 1286.

95. Claudia E. Haupt, *Licensing Knowledge*, 72 VAND. L. REV. 501, 540 (2019).

speaks in providing professional services – that supplies the theoretical justification for the imposition of liability on those who do not provide care in accordance with that community’s standard of care.<sup>96</sup> Thus, decision makers imposing liability for such speech do not do so on the basis of its *falseness* per se, but on its failure to convey accurately the expert consensus they purport to represent.

While this kind of speech regulation is essential to the preservation of trusted professions, it represents a narrow exception to the general principle that the government cannot sanction speech on the basis of disagreement with its content. The pandemic made clear just how narrow this exception could be. In 2022, California passed a law making it unprofessional conduct for a medical professional to “disseminate misinformation or disinformation related to COVID-19, including false or misleading information regarding the nature and risks of the virus, its prevention and treatment; and the development, safety, and effectiveness of COVID-19 vaccines.”<sup>97</sup> The law defined “misinformation” to include only treatment or advice about COVID-19 given to a patient under the physician’s care that was “contradicted by contemporary scientific consensus contrary to the standard of care.”<sup>98</sup>

When Governor Gavin Newsom signed the bill into law, he implicitly acknowledged the constitutional minefield created by efforts to regulate misinformation by insisting that this particular bill posed no problems because it was “narrowly tailored” to “egregious instances” and did not apply to “any speech outside of discussions directly related to COVID-19 treatment within a direct physician patient relationship.”<sup>99</sup> Therefore, he argued, while there may be legitimate reasons to be concerned about “the chilling effect other potential laws may have on [medical professionals],” this bill was different because the definition of misinformation was narrow and the law did not apply to public discourse.<sup>100</sup> One federal district court agreed, holding that the law was a permissible regulation of professional conduct.<sup>101</sup> Four weeks later, however, another federal district court found the law to be unconstitutionally vague, because “drawing a line between what is true and what is settled by scientific consensus is difficult, if not

---

96. *Id.* at 1241–42.

97. CAL. BUS. & PROF. CODE § 2270(a) (2022) (repealed 2023).

98. *Id.* § 2270(b)(4).

99. See Letter from Governor Gavin Newsom to Members of California State Assembly (Sept. 30, 2022), <https://www.gov.ca.gov/wp-content/uploads/2022/09/AB2098-signing-message.pdf> [<https://perma.cc/7Z7A-QBDM>].

100. See *id.*

101. McDonald v. Lawson, No. 22-cv-01805, 2022 WL 18145254, at \*11 (C.D. Cal. Dec. 28, 2022).

impossible.”<sup>102</sup> The court noted that this was all the more true in the context of COVID-19, “a disease that scientists have only been studying for a few years, and about which scientific conclusions have been hotly contested. COVID-19 is a quickly evolving area of science that in many aspects eludes consensus.”<sup>103</sup> Amidst continued litigation and controversy about the law’s constitutionality and chilling effects, California repealed the law in 2023.<sup>104</sup>

The California law was at once extremely narrow and problematically broad. It reached only medical advice given to an individual patient that was contrary to the standard of care. It is thus not even clear that the law would have enabled sanctions beyond existing restrictions on unprofessional conduct (although the law’s poor drafting made this ambiguous).<sup>105</sup> But by invoking the politicized notion of “misinformation” and giving it a broad and vague definition, the law raised fears of government overreach.<sup>106</sup> Even within the narrow bounds of a doctor-patient relationship, there is no such thing as “medical misinformation” writ large that the government could constitutionally target.

The California experience reflects the increasing reticence to treat even the speech of medical professionals as exempt from ordinary First Amendment rules prohibiting government regulation of falsity. The Supreme Court has encouraged this trend in recent years, in particular in its 2018 decision in *National Institute of Family & Life Advocates (NIFLA) v. Becerra*.<sup>107</sup> *NIFLA* made clear that the scope for regulation of physicians’ speech should be understood very narrowly lest it interfere with ordinary public discourse. *NIFLA* involved another California law, which required so-called “crisis pregnancy centers” to provide patients notices about state-provided free or low-cost reproductive health services and

---

102. Hoeg v. Newsom, 652 F. Supp. 3d 1172, 1190-91 (E.D. Cal. 2023).

103. *Id.* at 1188.

104. See Jacob Sullum, *California Quietly Repeals Restrictions on Doctors’ COVID-19 Advice*, REASON (Oct. 11, 2023, 11:00 AM), <https://reason.com/2023/10/11/california-quietly-repeals-restrictions-on-doctors-covid-19-advice> [https://perma.cc/8C76-WCDY]; Quinta Jurecic, *Good Luck Fighting Disinformation*, ATLANTIC (Mar. 26, 2024), <https://www.theatlantic.com/ideas/archive/2024/03/when-fighting-disinformation-doesnt-work/677879> [https://perma.cc/FA3R-99DT].

105. See, e.g., Kory v. Bonta, No. 24-cv-00001, 2024 WL 1742037, at \*9 (E.D. Cal. Apr. 23, 2024) (affirming the constitutionality of California medical boards’ powers to discipline physicians for conveying inaccurate COVID-19-related information to their patients and noting that “[i]t cannot be the case that scientific disagreement and inconsistencies in public health recommendations exempt doctors from the requirement that they adhere to the standard of care”).

106. See, e.g., Katherine Drabiak, *The Problem with Using Medical Boards to Regulate Misinformation*, 43 J. LEGAL MED. 3, 7 (2023).

107. Nat’l Inst. of Fam. & Life Advocs. (NIFLA) v. Becerra, 585 U.S. 755, 768 (2018) (“This Court’s precedents do not recognize such a tradition for a category called ‘professional speech.’”).

required unlicensed clinics to notify patients that they were not licensed.<sup>108</sup> In invalidating the law, Justice Thomas’s opinion for the Court dismissed the argument that the state had more leeway to regulate the speech of individuals in licensed professions more generally.<sup>109</sup> Such a rule would be dangerous, Justice Thomas insisted, because “[p]rofessionals [including doctors and nurses] might have a host of good-faith disagreements” and so (citing Justice Holmes in *Abrams*) “the best test of truth is the power of the thought to get itself accepted in the competition of the market.”<sup>110</sup> In Justice Thomas’s view, then, medical knowledge is, for First Amendment purposes, like any other knowledge, and the best form of regulation is the marketplace of ideas.<sup>111</sup>

Robert Post has remarked on the “breathtaking inanity of Justice Thomas’s invocation of the ‘marketplace of ideas’ in the context of professional speech.”<sup>112</sup> While there are contexts in which scholarly debate and the robust exchange of professional views is important, “[t]he law constructs such professional relationships to protect the reliance interests of patients and clients. It does not construct such relationships to embody the value of caveat emptor, as does the marketplace of ideas.”<sup>113</sup> It certainly seems somewhat fantastical to suggest that the remedy for false speech in a doctor’s exam room is simply “more speech.”

But acknowledging the particular context in which speech that is medical care takes place does not mean, and has never meant, that the medical profession cannot be wrong or that medical care cannot be politicized.<sup>114</sup> In this domain, as in any other, the power to declare orthodoxy and to punish dissent can be abused and can impede the advancement of knowledge. Scientific progress necessarily

---

108. *Id.* at 760–61.

109. *Id.* at 767.

110. *Id.* at 772 (quoting *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting)).

111. Justice Thomas did acknowledge, however, that medical-malpractice suits and informed-consent requirements – both of which are doctrinally distinct to medical practice – could be constitutional. *Id.* at 769–70; see also Haupt, *supra* note 81, at 188 (highlighting the *NIFLA* Court’s “insistence that it has never recognized professional speech as a category”).

112. Robert Post, *NIFLA and the Construction of Compelled Speech Doctrine*, 97 IND. L.J. 1071, 1083 n.68 (2022).

113. *Id.*

114. See, e.g., Haupt, *supra* note 87, at 1252–53 (noting that “the American Medical Association was at the forefront of the campaign to criminalize abortion in the nineteenth century, and the American Psychological Association [APA] did not declassify homosexuality as a mental disorder until 1973”); Timothy MacDonnell, *Disciplining Doctors: A Call for Caution when Responding to Physicians’ Counter-Consensus Speech in the Time of COVID-19*, 90 TENN. L. REV. 865, 929–30 (2023) (“Since the beginning of the pandemic, political figures have used doctors and scientists to advance their objectives.”).

requires the contestation and displacement of prevailing wisdom. There must be some competition of ideas, in other words, and the power to constrain that competition should be bestowed with great caution.

The case law contains many examples of the dangers of allowing the government to interfere with physicians' speech, given the potential politicization of medical care. The law at issue in *NIFLA* sought to facilitate access to reproductive health care, but in other states the government has used "informed consent" statutes to obstruct such access. Pennsylvania, for example, created a mandate that before performing an abortion, a doctor must inform the patient that there are state-published materials available that "describ[e] the fetus and provid[e] information about medical assistance for childbirth, information about child support from the father, and a list of agencies which provide adoption and other services as alternatives to abortion."<sup>115</sup> Meanwhile, Idaho sought to prevent medical providers in the state from providing patients with information about abortion services in other states, arguing that such information would not be protected speech but simply professional conduct.<sup>116</sup> And while California sought to punish doctors who knowingly gave their patients false information, other states introduced laws attempting to *prevent* state medical boards from disciplining doctors for spreading false information during the COVID-19 pandemic.<sup>117</sup> These few examples show that even this narrow exception to the general First Amendment rule against content-based regulation in the context of doctor-patient relationships is susceptible to politicization and governmental abuse – and that there are good reasons for keeping the exception narrow.

As these examples also suggest, despite the polarization of debates about medical misinformation in recent years, First Amendment protections seeking to keep the government out of regulating medical truth do not necessarily have a particular political valence. For example, in 2021 Democratic Senator Amy Klobuchar introduced a bill seeking to create a carveout from platforms' current statutory immunity for user-generated content for "health misinformation," where "health misinformation" would be in part defined by the Secretary of Health and Human Services.<sup>118</sup> Senator Klobuchar repeatedly denounced platforms' failure to take adequate action against false claims on their services, and repeatedly pointed to their failure to remove accounts belonging to the so-called

---

<sup>115</sup>. *Planned Parenthood of Se. Pa. v. Casey*, 505 U.S. 833, 881 (1992).

<sup>116</sup>. *Planned Parenthood Great Nw. v. Labrador*, No. 23-35518, 2024 WL 4966057, at \*12 (9th Cir. Dec. 4, 2024).

<sup>117</sup>. Rita Rubin, *When Physicians Spread Unscientific Information About COVID-19*, 327 JAMA 904, 906 (2022).

<sup>118</sup>. Health Misinformation Act of 2021, S. 2448, 117th Cong. (2021).

“Disinformation Dozen” as proof that they were not doing enough.<sup>119</sup> One of the Disinformation Dozen was, it turns out, Robert F. Kennedy, Jr. — Trump’s nominee for Secretary of Health and Human Services — whose views often conflict with the medical establishment but who would have been empowered under the terms of Klobuchar’s bill to define “health misinformation.” It is for this exact reason that Klobuchar’s bill would have been unconstitutional — because the bounds of public discourse and the meaning of “misinformation” should not be so susceptible to being fixed by a political actor.

Klobuchar’s bill never became law, however. And indeed, the result of the prevailing doctrinal architecture during the pandemic was that very few people were legally sanctioned for spreading medical misinformation. In part because of the limits that the First Amendment imposes on their power, government officials pressured platforms to do what they could not themselves do — remove medical misinformation from the public sphere. But this caused a different problem, because while platforms may have the technical means and legal latitude to police speech on their services,<sup>120</sup> they did not have the expertise or legitimacy to do so.

### III. PLATFORMS ARE NOT KNOWLEDGE COMMUNITIES

What the preceding potted summary of the First Amendment’s treatment of false or controversial medical claims suggests is that the legitimacy of suppressing speech depends on either a very direct and specific link to harm, or the existence of particular kinds of power inequalities and vulnerabilities. As this Part explains, it was the absence of these legitimating conditions that made platforms’ expanded content moderation efforts during COVID-19 so vulnerable to politicization. This does not mean that platforms have no role to play as gatekeepers. Some platforms have responded to the shifting political winds by publicly denouncing the expansiveness of their prior efforts, with little further

---

119. Press Release, Off. of Sen. Amy Klobuchar, Klobuchar Statement on Advisory from Surgeon General on Coronavirus Misinformation (July 15, 2021), <https://www.klobuchar.senate.gov/public/index.cfm/2021/7/klobuchar-statement-on-advisory-from-surgeon-general-on-coronavirus-misinformation> [https://perma.cc/T3A5-SHQ5].

120. This is not a claim that platforms could do this perfectly. Indeed, content moderation is impossible. Douek, *supra* note 17, at 791-92; Mike Masnick, *Masnick’s Impossibility Theorem: Content Moderation at Scale Is Impossible to Do Well*, TECHDIRT (Nov. 20, 2019, 9:31 AM), <https://www.techdirt.com/articles/20191111/23032743367/masnick-s-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well.shtml> [https://perma.cc/R5EL-KBXE]. But platforms can (and do) create rules and enforce them against content, even if that process will be error prone.

reflection or explanation of what their approach would be going forward.<sup>121</sup> But blanket renunciation of efforts to moderate medical misinformation in the face of political blowback suffers from the same problem that the aggressive policing of COVID-related misinformation did: it represents a political, rather than a principled, response to a public-health problem.

Instead, what is needed is an affirmative vision for the role of platforms in policing medical claims. This role may be more limited than the one platforms assumed during the pandemic, but it surely cannot be a complete abdication of responsibility. The path out of the current hyper-politicization requires more than just the admission of error, but also a full accounting of what happened, what worked, and what did not. As the only institutions with full access to the data showing what happened on their platforms during the pandemic, the social media companies are the only ones that can make this happen. Unfortunately, there are few incentives for platforms to facilitate this nuanced and realistic conversation. But this Part optimistically outlines what that conversation could look like, regardless.

To begin with, and to state what should be obvious, platforms do not occupy the same sociological position as the medical profession when it comes to the regulation of health-related information. Platforms' authority to regulate speech comes from their own status as private actors, rather than any special claim to expertise. Justice Kagan's opinion for the majority in *Moody v. NetChoice* expressly affirmed, as a passing example, a platform's right to "disfavor posts because they . . . discourage the use of vaccines."<sup>122</sup> But the Court was also clear that platforms would be protected if they made the opposite choice — lawmakers could not interfere with platforms' choices regardless of whether "the speech environment [lawmakers seek to create] is [better or] worse than the ones to which the major platforms aspire on their main feeds."<sup>123</sup> Thus, *Moody* is a decision in the same tradition as *Alvarez*. It affirms that platforms' content-moderation decisions are protected (like lies), not necessarily because they are valuable, good, or right, but because allowing the government to decide what or how speech should be disseminated is "a worse proposal" than leaving it to the private marketplace of ideas.<sup>124</sup>

This view of platforms as intermediaries is very different from how courts conceive of the role of the medical profession when it comes to the regulation of doctors' speech. As Part II explained, the First Amendment allows (and trusts) state-sanctioned professional organizations to impose certain limited content-

---

<sup>121.</sup> See *supra* notes 14–15.

<sup>122.</sup> 144 S. Ct. 2383, 2405 (2024).

<sup>123.</sup> *Id.*

<sup>124.</sup> *Id.* at 2403.

based restrictions on doctors' speech because of the specific characteristics of the speech and the professional community: namely, the fact that it is "individualized . . . , tied to a body of disciplinary knowledge from which it gains authority," and "occurs within a social relationship that is defined by knowledge asymmetry . . . and trust in the accuracy of that advice."<sup>125</sup> These features do not describe platforms' relationship with the speech on their services, with their users, or with broader society. Platforms have no special duty of care to their users, notwithstanding arguments that they should.<sup>126</sup> Nor do platforms have the authority of a knowledge community behind them when they write their rules.

Platforms dealt with this expertise deficit during the pandemic by insisting that they do not moderate speech on the basis of what *they* think is true or false but instead rely on the guidance of public-health authorities. Platforms still did not want to be "arbiters of truth," in other words. But the result was that platforms had no vocabulary or method for justifying decisions that departed from official guidance (for example, when medical consensus seems to have outpaced CDC guidance, as with masking guidelines), or for mediating between conflicting advice provided by different authorities. Professional institutions have shared methodologies and hierarchies of knowledge to guide their decision-making in such contexts, but platforms do not.<sup>127</sup> For all of these reasons, their interventions in discourse about COVID-19 lacked the legitimacy of the medical authorities that they sought to invoke.

At the same time, it remains true that online health misinformation is a serious public-health concern that costs lives.<sup>128</sup> Appreciating the particular role that social media platforms play in society does not mean they should throw up their hands—but it counsels caution and precision in what, exactly, platforms are asked to do. Assuming more content removal is always better, as some politicians

---

<sup>125</sup>. Haupt, *supra* note 95, at 529 (citations omitted).

<sup>126</sup>. See generally Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186 (2016) (arguing that online service providers such as social media platforms should be considered "information fiduciaries" that have "special duties to act in ways that do not harm the interests of the people whose information they collect, analyze, use, sell, and distribute"). But see generally Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2019) (taking a skeptical view of Jack M. Balkin's argument).

<sup>127</sup>. Claudia E. Haupt, *Unprofessional Advice*, 19 U. PA. J. CONST. L. 671, 705-10 (2017).

<sup>128</sup>. Michelle M. Mello, *Vaccine Misinformation and the First Amendment—The Price of Free Speech*, 3 JAMA HEALTH F. art. no. e220732, at 1 (2022) ("As many as 12 million persons may have forgone COVID-19 vaccination in the U.S. because of misinformation, resulting in an estimated 1200 excess hospitalizations and 300 deaths per day. If 5 fully loaded 747s crashed each week due to wrong information, regulators would be apoplectic.").

appear to,<sup>129</sup> ignores the risks and downsides involved in overly aggressive content moderation. Creating an affirmative vision for the role platforms should play requires reckoning with these risks, too.

The most obvious risk is that platforms, relying on authorities, will get it wrong and remove what is, with the benefit of hindsight, valuable information.<sup>130</sup> How heavily one weighs this risk will vary based on one's confidence in the judgment of particular medical authorities. But the risk of false positives is not the only risk created when platforms overstep. There are other downsides that would continue to exist, even if the false-positive problem could be reduced to zero.

A second, underappreciated risk arises from the function of platforms as dynamic expressive spaces and their role in trust formation. Social media platforms are a very particular kind of online space. They are not, for example, the same as online encyclopedias or medical reference sites. Instead, they are *social* spaces—people look to platforms to facilitate their relationships, rather than to provide them with particular information. This does not mean they are not important spaces of political and public discourse and education. Indeed, social media platforms are a significant source of news for many.<sup>131</sup> As the Supreme Court has recognized, they are among the “most important places” for the public to “celebrate some views, to protest others, or simply to learn and inquire.”<sup>132</sup> But they are also fundamentally messy spaces that contain a wide spectrum of human experience and cannot be made overly orderly. They facilitate democratic culture by allowing “ordinary people to participate freely in the spread of ideas and in the creation of meanings that, in turn, help constitute them as persons.”<sup>133</sup> This culture depends on people having access to these expressive spaces to engage in such public discourse, even when they are wrong. This process can be especially important in the context of a crisis. As crisis-informatics researcher Kate Starbird has explained, during times of stress, people “search for, disseminate, and

---

129. See, e.g., Shannon Bond, *Democrats Want to Hold Social Media Companies Responsible for Health Misinformation*, NPR (July 22, 2021, 3:59 PM ET), <https://www.npr.org/2021/07/22/1019346177/democrats-want-to-hold-social-media-companies-responsible-for-health-misinformation> [https://perma.cc/C6PH-E5CN]; Shannon Bond, *An Anti-Vaccine Book Tops Amazon's COVID Search Results. Lawmakers Call Foul*, NPR (Sept. 9, 2021, 4:44 PM ET), <https://www.npr.org/2021/09/09/1035559330/democrats-slam-amazon-for-promoting-false-covid-cures-and-anti-vaccine-claims> [https://perma.cc/K6TH-63L7].

130. See *supra* Part I for examples of this in the context of the pandemic.

131. *Social Media and News Fact Sheet*, PEW RSCH. CTR. (Sept. 17, 2024), <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet> [https://perma.cc/8SD7-T7SX].

132. *Packingham v. North Carolina*, 582 U.S. 98, 104 (2017).

133. Jack M. Balkin, Commentary, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 4 (2004).

synthesize the content [they] see into narratives” in a process called “collective sensemaking” that is “critical for [] decision-making, and in many cases allows us to relieve some of the anxiety and uncertainty we face in order to take action.”<sup>134</sup>

This messiness is important not only for people’s autonomy interest in self-expression, but also for the production of durable and trustworthy information. The production and dissemination of knowledge depends on openness and participation. And crucially, the robustness of truth depends on its ability to withstand dissent. The reliability of expert knowledge in fact *depends* on the ability of people to challenge prevailing wisdom, and for expert consensus nevertheless to remain unchanged.<sup>135</sup> This plays the important function of not only stress-testing expert consensus, but also fostering trust in that consensus because of its ability to be transparently contested. As sociologist Zeynep Tufekci has explained, “Misinformation is not something that can be overcome solely by spelling out facts just the right way. Defeating it requires earning and keeping the public’s trust.”<sup>136</sup> And this trust can be undermined, rather than fostered, by overly blunt content removals. As social media researcher and former pro-vaccine activist Renee DiResta has argued, “Social-media takedowns are not the right approach to addressing [misinformation like coronavirus-related scare-mongering] because they turn the propaganda into forbidden knowledge, often increasing the demand.”<sup>137</sup> Content-moderating away the supply of false claims does not make the demand for it disappear—indeed, it can have the opposite effect.

Thus, the idea that there is a linear relationship between more moderation of false claims and greater access and trust in truthful information relies on an overly static view of the information environment and a misconception of the social function of these expressive spaces. The opposite may often be true—

---

<sup>134</sup>. Kate Starbird, *Misinformation, Crisis, and Public Health – Reviewing the Literature*, MEDIAWELL (June 25, 2020), <https://mediawell.ssrc.org/research-reviews/misinformation-crisis-and-public-health> [https://perma.cc/E55J-QVFM].

<sup>135</sup>. Eugene Volokh, *When Are Lies Constitutionally Protected?*, 4 J. FREE SPEECH L. 685, 696 (2024).

<sup>136</sup>. Zeynep Tufekci, Opinion, *An Object Lesson from Covid on How to Destroy Public Trust*, N.Y. TIMES (June 8, 2024), <https://www.nytimes.com/2024/06/08/opinion/covid-fauci-hearings-health.html> [https://perma.cc/U8HM-B3DF]; see also Will Jennings, Gerry Stoker, Hannah Bunting, Viktor Orri Valgarosson, Jennifer Gaskell, Daniel Devine, Lawrence McKay & Melinda C. Mills, *Lack of Trust, Conspiracy Beliefs, and Social Media Use Predict COVID-19 Vaccine Hesitancy*, 9 VACCINES art. no. 593, at 10–11 (2021) (finding that trust in science and health organizations predicts willingness to take a vaccine).

<sup>137</sup>. Renée DiResta, *Anti-Vaxxers Think This Is Their Moment*, ATLANTIC (Dec. 20, 2020), <https://www.theatlantic.com/ideas/archive/2020/12/campaign-against-vaccines-already-under-way/617443> [https://perma.cc/E7KM-X4WZ].

people may develop distrust and suspicion in an environment they perceive as slanting the playing field. This is a problem not only because it undermines people's belief in trusted medical knowledge, but because it may undermine the legitimacy of core democratic processes. As Robert Post has put it:

In a democratic society, the revenge of the repressed can be a terrible thing. In dealing with the undoubted problem of misinformation, we must negotiate between the Scylla of widely circulating falsehoods and the Charybdis of the loss of democratic participation. Under conditions of polarization, suppression that is experienced as illegitimate can easily lead to an existential opposition between friends and enemies that would undermine the very possibility of democratic politics.<sup>138</sup>

These risks do not, however, mean that platforms can or should do nothing.

First, there is clearly still a role for content moderation, even if more limited than the approach platforms took during the pandemic. In limited circumstances where a particular post might have a direct relationship to serious physical harm, platforms might justifiably intervene. Posts advising people to use a harmful or ineffective cure for a deadly disease, for example, might be the kind of post a platform should remove, but—as Meta's own Oversight Board suggested—perhaps not in a context where the particular cure is not available and the post is made as part of political debate about whether it should be.<sup>139</sup> Meanwhile, claims that are less directly related to physical harm (for example, that COVID-19 is spread faster due to 5G networks or potentially false claims about the origin of the virus) should be left up. And for the same reason that First Amendment doctrine holds that false commercial speech is unprotected, platforms should have more stringent rules as to what they allow to be advertised on their sites. Free-speech concerns do not entirely disappear in the context of ads, but the role of platforms is very different when they are facilitating advertising as compared with when they are facilitating social interaction. They have a more legitimate claim to imposing their own standards on what kinds of economic transactions they facilitate.

Second, platforms can take other steps to increase the trust in and legitimacy of their content-moderation efforts. Platforms should implement structural separations between those charged with enforcing content-moderation rules and

---

<sup>138</sup>. Robert Post, *The Internet, Democracy and Misinformation*, in *DISINFORMATION, MISINFORMATION AND DEMOCRACY* (A. Koltay C. Garden & R. Krotoszynski eds., forthcoming 2024).

<sup>139</sup>. See, e.g., *Claimed Covid Cure*, OVERSIGHT BD. (Jan. 28, 2021), <https://www.oversight-board.com/decision/FB-XWJQBU9A> [<https://perma.cc/HG65-MAPC>] (overturning Facebook's decision to remove a post alleging a scandal relating to the availability of "hydroxychloroquine combined with azithromycin for use against COVID-19").

those interacting with government officials.<sup>140</sup> This will not eliminate concerns about illegitimate government influence, but it will mitigate them and help bolster trust that platforms are moderating fairly and consistently, rather than at the behest of political actors.

Finally, platforms should enable robust research into the online information ecosystem and what kinds of content-moderation interventions might be effective. Platforms sit on a wealth of data about people's information diets, how content flows online, and what kinds of platform interventions are effective, but a lack of access to the data for researchers or the public means that these insights – or the raw data that could lead to them – are not available to people outside the platforms themselves.<sup>141</sup> During the pandemic, platforms conducted what was in essence a worldwide social experiment about the effects of content-moderation interventions – from removing more content, to labeling false claims, to elevating authoritative information<sup>142</sup> – but years later there has been little transparency about the nature and extent of these interventions (beyond some high-level statistics),<sup>143</sup> let alone their impacts. Platforms' toolbox for responding to medical misinformation includes far more than simply deciding to take down speech or leave it up, as they showed during the pandemic, but some of these measures will be more effective than others. When Meta's own Oversight Board recommended that Meta conduct a review of the measures it took during the pandemic and publicly release the findings, Meta demurred and said that it thought its "resources are best deployed to prepare for future questions and crises, rather than attempt to conduct a broad and complex assessment based on a limited data set."<sup>144</sup> In a changed political environment, there is little appetite to

140. For a fuller exploration of this recommendation, see Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 586-89 (2022).

141. NATHANIEL PERSILY & JOSHUA A. TUCKER, SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM 320 (2020) ("[E]mployees of the platforms will be the only ones analyzing the data and learning the answers to the most pressing questions as to social media's impact on democracy and other social phenomena.").

142. Nandita Krishnan, Jiayan Gu, Rebekah Tromble & Lorian C. Abroms, *Research Note: Examining How Various Social Media Platforms Have Responded to COVID-19 Misinformation*, 2 HARV. KENNEDY SCH. MISINFORMATION REV. (Dec. 15, 2021), <https://misinforeview.hks.harvard.edu/article/research-note-examining-how-various-social-media-platforms-have-responded-to-covid-19-misinformation> [<https://perma.cc/9NC8-TKUH>].

143. See, e.g., Neal Mohan, *Perspective: Tackling Misinformation on YouTube*, YOUTUBE OFF. BLOG (Aug. 25, 2021), <https://blog.youtube/inside-youtube/tackling-misinfo> [<https://perma.cc/KBB8-QUEJ>]; *Removal of COVID-19 Misinformation*, OVERSIGHT BD. (Apr. 20, 2023), <https://www.oversightboard.com/decision/pao-sabu4p2s> [<https://perma.cc/66LE-9A7S>].

144. *Meta Q3 2023 Quarterly Update on the Oversight Board*, META TRANSPARENCY CTR. 24 (Nov. 2023), <https://transparency.meta.com/oversight/meta-biannual-updates-on-the-oversight-board> [<https://perma.cc/JYH9-G3TT>].

talk about and learn from the measures that platforms took during the pandemic. But limited outside access to data has hampered any rigorous and independent analysis of the extraordinary measures taken during this time of crisis.<sup>145</sup>

Such research can result in counterintuitive findings. Early literature on social media platforms focused on fears of echo chambers and filter bubbles in which algorithms fed them an information diet that simply reinforced their existing views,<sup>146</sup> but there is now a growing consensus that such fears have been overblown.<sup>147</sup> Similarly, “[w]here once it was a common concern that retractions may backfire and people may believe even more in the misinformation after the correction is presented, recent research has found this phenomenon to be rare.”<sup>148</sup> As these few examples show, the science on the best way to facilitate trust and belief in factual information is evolving—and will continue to do so as the online environment itself changes—but people’s intuitions about the problems and their solutions are not always correct.

In better moments, public discourse about content moderation of misinformation acknowledges all these nuances. When Vivek H. Murthy, the Surgeon General of the United States (and one not shy about his concerns about the health impacts of social media<sup>149</sup>) issued an advisory on “Confronting Health Misinformation,” perhaps cognizant of First Amendment concerns with government demands for more censorship, he stopped short of calling for platforms to remove more content.<sup>150</sup> Instead, he called for platforms to take a number of

---

145. Rory Smith et al., *A Systematic Review Of COVID-19 Misinformation Interventions: Lessons Learned*, 42 HEALTH AFFS. 1738, 1739 (2023).

146. See generally, e.g., CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA (2017) (describing the effects of social media echo chambers).

147. Briony Swire-Thompson & David Lazer, *Public Health and Online Misinformation: Challenges and Recommendations*, 41 ANN. REV. PUB. HEALTH 433, 438 (2020); Andrew Guess, Benjamin Lyons, Brendan Nyhan & Jason Reifler, *Avoiding the Echo Chamber About Echo Chambers: Why Selective Exposure to Like-Minded Political News Is Less Prevalent than You Think*, KNIGHT FOUND. 3 (2018), [https://kf-site-production.s3.amazonaws.com/media\\_elements/files/000/000/133/original/Topos\\_KF\\_White-Paper\\_Nyhan\\_V1.pdf](https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf) [<https://perma.cc/8X2A-8K4E>].

148. Swire-Thompson & Lazer, *supra* note 147, at 445 (citing Briony Swire, Adam J. Berinsky, Stephan Lewandowsky & Ullrich K.H. Ecker, *Processing Political Misinformation: Comprehending the Trump Phenomenon*, 4 ROYAL SOC’Y OPEN SCI. art. no. 160802, at 1 (2017); Briony Swire, Ullrich K.H. Ecker & Stephan Lewandowsky, *The Role of Familiarity in Correcting Inaccurate Information*, 43 J. EXPERIMENTAL PSYCH.: LEARNING, MEMORY & COGNITION 1948, 1948 (2017); Thomas Wood & Ethan Porter, *The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence*, 41 POL. BEHAV. 135, 135 (2018).

149. Vivek H. Murthy, Opinion, *Surgeon General: Why I’m Calling for a Warning Label on Social Media Platforms*, N.Y. TIMES (June 17, 2024), <https://www.nytimes.com/2024/06/17/opinion/social-media-health-warning.html> [<https://perma.cc/W4NA-QSXV>].

150. Murthy, *supra* note 32, at 12.

other measures, including giving researchers access to data to analyze the spread of misinformation, addressing information deficits, amplifying communications from trusted messengers, and protecting health professionals from online harassment.<sup>151</sup>

These kinds of careful claims do not make good soundbites, however, and political rhetoric often paints a much simpler (and, ironically, misleading) picture — a picture that is based on widely shared confusion about the role that platforms can and should play as intermediaries and gatekeepers. The politicization and backlash against platforms' content moderation during the pandemic is a predictable consequence of asking platforms to play a role they are not cut out for. It suggests that pushing for even more heavy-handed responses to medical misinformation may result in short-term wins (i.e., fewer false claims on particular services), but may undermine the longer-term goal of facilitating the processes of trust building and truth formation.

## CONCLUSION

Debates about content moderation have become so polarized and politicized that they admit no nuance. There are important lessons to be learned from the experience of moderating medical misinformation during the COVID-19 pandemic, but the political noise has drowned out measured conversation about what happened and what should happen going forward. Contrary to one side of the debate, platforms do have a role to play in the regulation of false information about COVID-19 and other health issues. But that role is, and has to be, a far more limited one than many on the other side of the political debate envision. It can be true both that online medical misinformation is a serious public-health problem and that excessive content moderation will exacerbate, rather than cure, the underlying disease of distrust.

*Assistant Professor, Stanford Law School. Many thanks to Ros Dixon, Claudia Haupt, Genevieve Lakier, Michelle Mello, and Yoel Roth for generous comments on earlier drafts, and to the wonderful editors of the Yale Law Journal for their care and hard work.*

---

<sup>151</sup>. *Id.*