TAMING ONLINE PUBLIC HEALTH MISINFORMATION

IRA RUBINSTEIN* & TOMER KENNETH†

The COVID-19 pandemic was shaped by a corollary infodemic: an abundance of public health misinformation ("PHM"), primarily online. Online PHM has pervasive effects, creating health hazards for individuals and hindering society's attempts to confront diseases and health risks. Troublingly, online PHM is a difficult problem to solve. It involves regulation of online speech, content moderation, First Amendment issues, and public health law. And like other regulations of misinformation, it raises intricate epistemic and normative questions. This Article discusses the problems associated with online PHM, points to shortcomings in existing responses, and advances two primary solutions. The Article contributes to existing scholarship by developing a comprehensive plan for confronting online PHM, thereby also casting new light on online speech regulation.

The Article begins by developing the concept of PHM and discussing the major harms it poses, using COVID-19 as a main example. Next, it surveys how major platforms confronted online PHM during the COVID-19 pandemic and explains the shortcomings of relying on platforms to confront PHM. The Article then critiques existing regulatory measures that governments use to confront online PHM. Positively, the Article promotes two promising paths for confronting online PHM. First, soft-regulation measures—specifically voluntary self-regulation and voluntary enforcement. Such approaches were successfully implemented around the world to confront online speech harms, but so far mostly overlooked in the U.S. Second, it explores a new approach to managing online speech: regulating algorithmic recommendation (and amplification). Drawing on a technical primer, recent bills, and caselaw, the Article argues—contrary to popular views—that regulation of algorithmic recommendation can survive First Amendment scrutiny.

* Senior Fellow, Information Law Institute, New York University School of Law.

†JSD Candidate; Fellow, Information Law Institute, New York University School of Law. We thank Kathy Strandburg, Ari Waldman, Alexandre de Streel, Daniel Schwarcz, Ken Shear, Caroline Mala Corbin, and the fellows at NYU's Information Law Institute for discussions about earlier versions. Many thanks also to participants in the Yale Law School 10th Freedom of Expression Scholars Conference and the 15th Privacy Law Scholars Conference (PLSC, Northeastern University) for helpful comments. We are also grateful to Sergi Galvez Duran, Justin Jin, and Sharngan Aravindakshan for excellent research assistance. Finally, we sincerely thank the *Harvard Journal on Legislation* editors for exceptionally thoughtful and constructive suggestions, which helped improve the Article.

I. Introduction

As COVID-19 was causing sickness and death in unprecedented numbers, people were searching for ways to confront this horrible disease.[1] One scientific study found that an FDA-approved drug might reduce the viral load of COVID-19 and suggested further investigation.[2] News about this new "treatment" caught fire, especially on social media platforms.[3] Soon enough, hundreds of thousands of people sought and obtained the drug—ivermectin—and its use skyrocketed.[4] There was only one problem: ivermectin's intended use was to treat parasitic worms in humans and ani-

[1] *See, e.g.*, Jon Cohen & Kai Kupferschmidt, *The 'Very, Very Bad Look' of Remdesivir, the First FDA-Approved COVID-19 Drug*, Science (Oct. 28, 2020), https://www.science.org/content/article/very-very-bad-look-remdesivir-first-fda-approved-COVID-19-drug [https://perma.cc/6J4D-8CXM] (detailing discussions about treatments like hydroxychloroquine, "convalescent" plasma, and remdesivir).

[2] Leon Caly, Julian D. Druce, Mike G. Catton, David A. Jans & Kyle M. Wagstaff, *The FDA-Approved Drug Ivermectin Inhibits the Replication of SARS-CoV-2 in Vitro*, 178 Antiviral Rsch. 104787, 104787 (2020).

[3] Davey Alba, *Facebook Groups Promoting Ivermectin as a COVID-19 Treatment Continue to Flourish.*, N.Y. Times (Sept. 28, 2021), https://www.nytimes.com/2021/09/28/technology/facebook-ivermectin-coronavirus-misinformation.html [https://perma.cc/646K-25NE].

[4] Emma Goldberg, *Demand Surges for Deworming Drug for Covid, Despite Scant Evidence It Works*, N.Y. Times (Aug. 30, 2021), https://www.nytimes.com/2021/08/30/health/covid-ivermectin-prescriptions.html [https://perma.cc/T8FC-NKPM].

mals, and it was neither effective nor safe for treating COVID-19.[5] Major social media platforms responded quickly by flagging false content about the drug, directing users to accurate sources of information, and deleting groups dedicated to distributing the drug.[6] Public health agencies also voiced concerns, emphasizing the risks of ivermectin and calling on the public not to use it against COVID-19.[7] And yet, misinformation about ivermectin persisted, alongside misinformation celebrating other "miracle drugs" and undermining the efficacy and safety of the real vaccines.[8] The spread of online misinformation about COVID-19 and unproven treatments resulted in a serious public health problem and individual suffering. Unfortunately, the ivermectin story—originating from a misunderstanding of science, propagated and amplified through social media, surviving platforms' and officials' mitigation efforts, and risking people's health—is far from unique.[9] It illustrates the tangible harms and complex challenges posed by online public health misinformation ("PHM").

This Article explores online PHM: what it is and how to confront it. Most legal studies about this topic adopt a narrow perspective, emphasizing one solution—more or less innovative—or another. However, confronting online PHM is challenging because PHM—like other forms of online misinformation—is a multidimensional problem. A comprehensive solution must consider various perspectives from different legal fields. This Article methodically explores the major legal aspects of online PHM: the complex nature of misinformation, public health law, consumer protection and tort law, cooperation between governments and private platforms, and First Amendment implications of regulating algorithmic amplification of content on social media platforms. By exploring a wide range of solutions, some new and some improvements of existing approaches, the Article brings all those issues into dialogue with one another and thereby provides a clear and comprehensive analysis of the possible legal responses to online PHM.

This Article supports two positive interventions for confronting online public health misinformation. First, several soft regulation measures have been quite successful outside the United States and could be adopted here as well.[10] Soft regulation measures include guidelines, codes of conduct, coop-

---

[5] *Why You Should Not Use Ivermectin to Treat or Prevent COVID-19*, FDA (Dec. 10, 2021), https://www.fda.gov/consumers/consumer-updates/why-you-should-not-use-ivermectin-treat-or-prevent-covid-19 [https://perma.cc/47DV-3M8X].

[6] *See infra* Part III; *see also* Dawn Carla Nunziato, *Misinformation Mayhem: Social Media Platforms' Efforts to Combat Medical and Political Misinformation*, 19 First Amend. L. Rev. 32, 37–51 (2020).

[7] *See Why You Should Not Use Ivermectin to Treat or Prevent COVID-19*, *supra* note 5.

[8] Alba, *supra* note 3. For a survey of COVID-19 misinformation, see *infra* Part II.B.

[9] *See* Joan Donovan, Jennifer Nilsen, Gabrielle Lim, Nikhil George, Danielle Levin & Jessica Leon, *Trading Up the Chain: The Hydroxychloroquine Rumor*, Media Manipulation Casebook (Mar. 5, 2021), https://mediamanipulation.org/case-studies/trading-chain-hydroxychloroquine-rumor [https://perma.cc/JA9H-4UHV] (discussing a similar story with regard to hydroxychloroquine as a cure for COVID-19).

[10] *See infra* Part V.A.

eration, and other nonbinding instruments that states use to influence private actors that are "weaker" than traditional laws and regulations.[11] These soft regulations, we argue, can help harness the powers of social media platforms for governing online speech. These mechanisms are required because of the limitations of social media self-regulation and the shortcomings of existing legal tools.[12] Second, we discuss a path for traditional regulation of social media to confront PHM: regulation of algorithmic recommendation and amplification. Although many commentators believe the First Amendment (and Section 230) hem in government action against misinformation, we defend regulation of algorithmic recommendation and amplification. We distinguish algorithmic amplification from content moderation based on their technical differences, and explain the legal implications of these distinctions.[13] Our analysis paves the way to regulating recommendation algorithms, primarily through adding friction and middleware.

The Article addresses the problem of regulating online misinformation more generally. Online PHM is a particularly potent case study for researching online misinformation. PHM raises all the thorny known problems of misinformation on social media, including complicated questions of content moderation and platform regulation. Online PHM has two distinct features that helpfully narrow the scope of online misinformation: (1) the harm it poses is clear and demonstrable;[14] and (2) it is easier (though not unproblematic) to circumscribe the topic and distinguish between information and misinformation.[15] By analyzing the more contained problem of online PHM, we learn valuable lessons about more complicated issues such as political misinformation. Thus, our analysis of online PHM contributes to a pressing social and political question: how should states confront misinformation on social media?

The Article begins, in Part II, by explaining the notion of "public health misinformation," illustrating the grave problems it poses for individuals and societies, and situating online PHM within the broader context of misinformation. It will illustrate the need to discuss online PHM and the benefits such discussion can hold for regulating misinformation more generally. Part III considers how major social media platforms, like Facebook, YouTube, and Twitter (hereinafter "platforms" unless explicitly singling out one of them), confronted PHM related to COVID-19. The Part discusses the problems of relying only on platforms to confront online PHM, and argues in favor of governmental involvement. Part IV explains why the existing governmental regulatory responses to online PHM are lacking. Then, Part V develops positive arguments for the role governments can, and should, play

---

[11] *Soft Law,* OECD, https://www.oecd.org/gov/regulatory-policy/irc10.htm [https://perma.cc/R4WP-VFW7].

[12] *See infra* Part III and IV, respectively.

[13] *See infra* Part V.B.

[14] *See infra* Part II.B.

[15] *See infra* Part II.A.

in taming online PHM. It explores what governments can achieve using soft law measures that influence platforms to self-regulate and to enforce their policies in a manner conducive to confronting PHM. It suggests intensifying the use of government speech to confront online PHM. Finally, it analyzes new laws and regulations that try to confront online PHM and argues, against prevalent views, that regulation of algorithmic amplification can survive First Amendment limitations.

II. Public Health Misinformation

A. The Scope of Discussion: Online Public Health Misinformation

There is much that epidemiologists and the vaccine-hesitant[16] agree upon. They usually agree that freedom and health are important and desirable for individuals and communities and that balancing them is necessary. They probably also agree that inaccurate information about health-related issues is socially undesirable. That is, both groups are concerned that such inaccuracies would lead individuals and communities to decisions that may result in physical harms. However, epidemiologists and the vaccine-hesitant disagree, regularly and fiercely, about various factual health claims. For instance, they disagree about the safety of vaccines, the dangers and prevalence of some diseases, and the efficacy of certain treatments.[17] They disagree, that is, about what should count as information and misinformation.

By misinformation, we mean disseminated or propagated information that is false, regardless of the speaker's intention.[18] Concerns about the rise of fake news and the role of innovative technologies in spreading falsehoods have been around for at least a century.[19] But misinformation gained much of its vigor only in recent years as a specific notion and as part of a larger social phenomenon associated with disinformation, post-truth, and fake

---

[16] On the concept of vaccine hesitancy, see generally Noni E. MacDonald, SAGE Working Group on Vaccine Hesitancy, *Vaccine Hesitancy: Definition, Scope and Determinants*, 33 Vaccine 4161 (2015) (describing vaccine hesitancy and the factors that influence it). The World Health Organization listed vaccine hesitancy in the top ten global health threats for 2019. *Ten Threats to Global Health in 2019*, WHO, https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [https://perma.cc/G6U4-U5EX].

[17] *See* Cailin O'Connor & James Owen Weatherall, The Misinformation Age: How False Beliefs Spread 142–43 (2019); Dorit Rubinstein Reiss & John Diamond, *Measles and Misrepresentation in Minnesota: Can There Be Liability for Anti-Vaccine Misinformation That Causes Bodily Harm?*, 56 San Diego L. Rev. 531, 544–53 (2019).

[18] On this broadly accepted view of misinformation, see, e.g., Ben Epstein, *Why It Is So Difficult to Regulate Disinformation Online*, *in* The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States 190, 192 (Steven Livingston & W. Lance Bennett eds., 2020).

[19] *See generally* Edward McKernon, *Fake News and the Public*, Harper's Mag., Oct. 1925, at 528 (discussing the rise of fake news as early as 1921).

news.[20] Distinguishing between information and misinformation is often problematic because it raises complicated epistemic questions. These include first-order questions about the epistemic accuracy of specific truth-claims, such as how many people attended Trump's inauguration?[21] Does the MMR vaccine cause autism?[22] Does smoking cause cancer?[23] It also includes second-order questions, such as which experts should be trusted, or which methods are valid for determining what is true.[24]

In this Article, however, our specific focus on health-related claims simplifies many of these complications. We consider such claims valid or verified by adhering to the relevant science.[25] That is, for health-related factual claims, we distinguish information from misinformation according to the scientifically best understanding of the facts at a given time. Factual claims that align with the best available scientific understanding are regarded as information, and those that do not are regarded as misinformation. There are good reasons to support adherence to science at least with regards to health-related claims.[26] And democracies have a well-established history of adopting science as a lodestar.[27]

Admittedly, this approach does not solve all problems with misinformation, and many epistemic questions linger. For one, identifying a consensus within the scientific community regarding a specific claim is sometimes difficult.[28] And the scientific understanding itself is likely to change (and hopefully advance) over time.[29] For another, this approach invites second-order

---

[20] *See, e.g.*, Johan Farkas & Jannick Schou, *Prophecies of Post-Truth*, *in* POST-TRUTH, FAKE NEWS AND DEMOCRACY: MAPPING THE POLITICS OF FALSEHOOD 45 (2019); Symposium, *Falsehoods, Fake News, and the First Amendment*, 71 OKLA. L. REV. 1 (2019).

[21] Brian F. Schaffner & Samantha Luks, *Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us About the Source of Political Misinformation in Surveys*, 82 PUB. OP. Q. 135, 136–37 (2018).

[22] It does not. *See, e.g.*, *Autism and Vaccines*, CDC (Dec. 1, 2021), https://www.cdc.gov/vaccinesafety/concerns/autism.html [https://perma.cc/YL58-PH9F].

[23] It does. *See, e.g.*, *What are the Risk Factors for Lung Cancer?*, CDC (Oct. 25, 2022), https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm [https://perma.cc/LDL8-Q242].

[24] *See, e.g.*, Jo Fox, *'Fake News'—The Perfect Storm: Historical Perspectives*, 93 HIST. RSCH. 172, 182 (2020); Farkas & Schou, *supra* note 20; Jeroen de Ridder, *Deep Disagreements and Political Polarization*, *in* POLITICAL EPISTEMOLOGY 226 (Elizabeth Edenberg & Michael Hannon eds., 2021).

[25] *See generally* Wen-Ying Sylvia Chou, April Oh & William M. P. Klein, *Addressing Health-Related Misinformation on Social Media*, 320 JAMA 2417 (2018) (defining health misinformation).

[26] *See, e.g.*, NAOMI ORESKES, WHY TRUST SCIENCE? 55–59 (2021); O'CONNOR & WEATHERALL, *supra* note 17, at 44.

[27] *See generally* SOPHIA A. ROSENFELD, DEMOCRACY AND TRUTH: A SHORT HISTORY (2019).

[28] *See e.g.*, Emily K. Vraga & Leticia Bode, *Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation*, 37 POL. COMMC'N 136 (2020) (discussing scientific consensus with regard to health misinformation); Boaz Miller, *The Social Epistemology of Consensus and Dissent*, *in* THE ROUTLEDGE HANDBOOK OF SOCIAL EPISTEMOLOGY 230 (Miranda Fricker et al. eds., 2019) (same).

[29] Briony Swire-Thompson & David Lazer, *Public Health and Online Misinformation: Challenges and Recommendations*, 41 ANN. REV. PUB. HEALTH 433, 434 (2020).

discussions about which claims count as scientific consensus (or disagreement) and which are beyond its boundaries. These questions matter. As Claudia Haupt argues, it makes a difference, for legal analysis and outcomes, whether an alleged expert or professional disagrees with the scientific-medical consensus based on agreed-upon scientific validation processes (internal outlier), or based on reasons that are exogenous to the medical knowledge community (external outlier).[30] This fascinating endeavor to devise a "constitutional sociology of knowledge" is beyond the scope of this Article.[31] However, the existence of such borderline cases should not deter us. We continue under a practical and non-skeptical assumption that science exists, and that more often than not it can reliably answer health related questions.

In what follows, we take it for granted that science is what distinguishes health information from health misinformation. So, we define public health misinformation as disseminated or propagated health-related factual claims that are scientifically false.[32] Like most information (and misinformation) nowadays, health-related information is propagated and disseminated mostly on social media platforms, the "modern public square."[33] Accordingly, this paper focuses primarily on *online PHM*—especially on PHM that spreads using social media platforms.

Online PHM is highly problematic because it harnesses all the powers of communication via large platforms. Platforms allow almost real-time online communication between individuals and communities, overcoming space and time limitations at practically zero costs to speakers.[34] In addition, online PHM enjoys unprecedented velocity. A single PHM video can be quickly seen, shared, or otherwise engaged with by tens of millions of people around the world.[35] Considering that roughly 230 million Americans use

---

[30] *See* Claudia E. Haupt, *Unprofessional Advice*, 19 U. PA. J. CONST. L. 671, 690–91 (2017).

[31] *See* ROBERT C. POST, DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM 55–60 (2012); *see also* FREDERICK F. SCHAUER, THE PROOF: USES OF EVIDENCE IN LAW, POLITICS, AND EVERYTHING ELSE 15–54, 161–62 (2022) (explaining that the decision to trust science, for instance over astrology, is a sociological decision).

[32] For similar definitions, see OFFICE OF THE SURGEON GENERAL, CONFRONTING HEALTH MISINFORMATION: THE U.S. SURGEON GENERAL'S ADVISORY ON BUILDING A HEALTHY INFORMATION ENVIRONMENT 4 (2021) [hereinafter SG REPORT] (defining health misinformation as "information that is false, inaccurate, or misleading according to the best available evidence at the time").

[33] Packingham v. North Carolina, 582 U.S. 98, 107 (2017).

[34] *See generally* Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L.J. 1805 (1995).

[35] *See, e.g.*, Ellen P. Goodman & Karen Kornbluh, *Social Media Platforms Need to Flatten the Curve of Dangerous Misinformation*, SLATE MAGAZINE (Aug. 21, 2020), https://slate.com/technology/2020/08/facebook-twitter-youtube-misinformation-virality-speed-bump.html [https://perma.cc/SVK4-8KBG] (explaining how one PHM video gained 20 million views on Facebook alone within 12 hours).

platforms[36]—81% use YouTube regularly and 69% use Facebook at least daily[37]—such velocity translates to an impressive potential reach. Moreover, various actors can use platforms to target specific messages to specific (clusters of) individuals, as anti-vaccine groups who disseminate online PHM know too well.[38]

Despite public health officials' attempts to spread helpful preventive information on social media, platforms are infested with PHM.[39] This is hardly surprising given human nature. As Jonathan Swift wrote in 1710, and recent empirical analysis has vindicated, "[f]alsehood flies, and the truth comes limping after it."[40] Individuals and groups—whose motivations are quite varied—have been spreading PHM on social media and influencing people's health choices with relative ease, even before the COVID-19 pandemic.[41] Researchers found that PHM was widespread in social media discussions about MMR vaccines, as well as the Zika and Ebola viruses.[42] Experts have recognized this trend for a while, and warned that online PHM is a "global public-health threat" even before the COVID-19 pandemic.[43] Hence, the Congressional Research Service found that PHM "could be detrimental to public health and make efforts to address the pandemic or achieve public acceptance of a vaccination more challenging."[44]

Online PHM is particularly troubling in view of today's public health emergencies, as more people turn to social media to seek information on how to behave in response to evolving threats.[45] And indeed, online PHM

---

[36] Jason A. Gallo & Clare Y. Cho, *Social Media: Misinformation and Content Moderation Issues for Congress*, CONG. RSCH. SERV. 4–6 (Jan. 27, 2021), https://crsreports.congress.gov/product/pdf/R/R46662 [https://perma.cc/RT2C-P4SP].

[37] Brooke Auxier & Monica Anderson, *Social Media Use in 2021*, PEW RESEARCH CENTER: INTERNET, SCIENCE & TECH (April 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [https://perma.cc/M983-FLG2].

[38] *See infra* Part IV.C; Renée DiResta, *Anti-Vaxxers Think This Is Their Moment*, ATLANTIC (December 20, 2020), https://www.theatlantic.com/ideas/archive/2020/12/campaign-against-vaccines-already-under-way/617443/ [https://perma.cc/4QA3-YDB9].

[39] *See* Kenny Mendoza-Herrera, Isabel Valero-Morales, Maria E. Ocampo-Granados, Hortensia Reyes-Morales, Fernanda Arce-Amaré & Simón Barquera, *An Overview of Social Media Use in the Field of Public Health Nutrition: Benefits, Scope, Limitations, and a Latin American Experience*, 17 PREVENTING CHRONIC DISEASE E76, 1–3 (2020).

[40] Jonathan Swift, *Political Lying*, *in* 3 ENGLISH PROSE (Henry Craik ed., 1916); *see also* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE 1146, 1146 (2018).

[41] *See* Salman Bin Naeem & Maged N. Kamel Boulos, *COVID-19 Misinformation Online and Health Literacy: A Brief Overview*, 18 INT'L J. ENV'T RSCH. & PUB. HEALTH 8091, 8094 (2021).

[42] Yuxi Wang, Martin McKee, Aleksandra Torbica & David Stuckler, *Systematic Literature Review on the Spread of Health-Related Misinformation on Social Media*, 240 SOC. SCI. & MED. 112552, 112555 (2019).

[43] Heidi J. Larson, *The Biggest Pandemic Risk? Viral Misinformation*, 562 NATURE 309, 309 (2018). *See generally* Heidi J. Larson, *Blocking Information on COVID-19 Can Fuel the Spread of Misinformation*, 580 NATURE 306 (2020).

[44] Gallo & Cho, *supra* note 36, at 14–16.

[45] *Id.*

has peaked during the COVID-19 pandemic.[46] From the outset, the COVID-19 pandemic came hand-in-hand with the COVID-19 infodemic—the abundance of online PHM that caused confusion, undermined public-health efforts and drew significant attention from officials.[47] As the Surgeon General recently noted: while PHM is not new, "the speed, scale, and sophistication with which misinformation has been spread during the COVID-19 pandemic has been unprecedented."[48]

B. The Harms of Public Health Misinformation

Online PHM is prevalent and concerns about it are widespread. Are the concerns justified? What are the actual harms of online PHM? Admittedly, it is difficult to establish a firm causal connection between PHM and individuals' health decisions or inferior public health outcomes.[49] Research also suggests that sharing content on social media does not necessarily indicate that the sharer thinks the content is accurate.[50] So, perhaps we should just ignore online PHM as another rhetorical hyperbole of social media.[51] We disagree. To explain our position, we survey some of the actual harms of online PHM.

Individuals and societies make choices about how to lead their lives based on the information available to them.[52] Health-related information—including the potential dangers of some new virus, the dangers and benefits of some vitamin or supplement, what activities might lead one to become infected or infect others, and whether one can trust the safety and efficacy of a new vaccine—affects individuals' and societies' choices about their health.

---

[46] Vosoughi et al., *supra* note 40, at 1146.

[47] *See, e.g.*, John Zarocostas, *How to Fight an Infodemic*, 395 LANCET 676, 676 (2020) (noting the proliferation of the word "infodemic" to describe the spread of misinformation); *Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Tackling COVID-19 Disinformation - Getting the Facts Right*, COM (2020) 8 final (Oct. 6, 2020) [hereinafter *EU Joint Communication*] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEX:52020JC0008 [https://perma.cc/9R87-REES] (discussing the rise of the infodemic, including dangerous hoaxes and misleading healthcare information); *Infodemic,* WHO, https://www.who.int/health-topics/infodemic#tab=tab_1 [https://perma.cc/8JA6-BNCK] ("An infodemic is too much information including false or misleading information in digital and physical environments during a disease outbreak.").

[48] Impact of Health Misinformation in the Digital Information Environment in the United States Throughout the COVID-19 Pandemic Request for Information (RFI), 87 FED. REG. 12712, 12713 (Mar. 7, 2022), https://www.govinfo.gov/content/pkg/FR-2022-03-07/pdf/2022-04777.pdf [https://perma.cc/9K8M-RUGZ].

[49] *See, e.g.*, Wendy E. Parmet & Jason A. Smith, *Free Speech and Public Health: A Population-Based Approach to the First Amendment Symposium: Food Marketing to Children and the Law*, 39 LOY. L.A. L. REV. 363, 376–77 (2006).

[50] *See, e.g.*, Gordon Pennycook & David G. Rand, *The Psychology of Fake News*, 25 TRENDS COGNITIVE SCIS. 388, 391 (2021).

[51] *See, e.g.*, Clifford v. Trump, 339 F. Supp. 3d 915, 926 (C.D. Cal. 2018), *aff'd*, 818 F. App'x 746 (9th Cir. 2020) (analyzing tweets as rhetorical hyperbole that raise no legal liability).

[52] *See generally* Hugh G. Petrie, *Practical Reasoning: Some Examples*, 4 PHIL. & RHETORIC 29 (1971).

Obviously, these choices have important consequences: they mark the difference between healthy lives and illness or death, and often determine whether people can engage in meaningful social activities or must refrain from them. Moreover, as the COVID-19 pandemic makes evident, often one's health choices affect not only oneself, but one's entire community and its economic and social life.[53] Lawrence O. Gostin and Lindsay F. Wiley express it well:

> Health is foundationally important because of its intrinsic value and singular contribution to human functioning . . . . Physical and mental health allow individuals to recreate, socialize, work, and engage in family and social activities that bring meaning and happiness to their lives . . . . Health is also essential for the functioning of populations. Without minimum levels of health, people cannot fully engage in social interactions, participate in the political process, exercise rights of citizenship, generate wealth, create art, or provide for the common security. A safe and healthy population provides the basis for a country's government structures, social organizations, cultural endowment, economic prosperity, and national defense. Population health is a transcendent value because a certain level of human functioning is a prerequisite for activities that are critical to the public's welfare—social, political, and economic.[54]

Therefore, even if those who share PHM are not affected, sharing itself causes harm. Sharing online PHM has negative externalities because of the "illusory truth" effect, which makes repeated claims more likely to be judged as true. Hence, PHM persists despite contradictory advice from accurate sources, which in turn undermines the efficacy of future public health interventions.[55] Indeed, there is ample evidence supporting the contribution of PHM to worse health choices by individuals and communities.[56] For in-

---

[53] *See generally* JAMES G. HODGE, JR., PUBLIC HEALTH LAW IN A NUTSHELL (2021) (explaining that health is fundamental to, and affected by, all policies and social or economic activities).

[54] LAWRENCE O. GOSTIN & LINDSAY F. WILEY, PUBLIC HEALTH LAW: POWER, DUTY, RESTRAINT 7–8 (2016).

[55] Ullrich K. H. Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga & Michelle A. Amazeen, *The Psychological Drivers of Misinformation Belief and Its Resistance to Correction*, 1 NATURE REVS. PSYCH. 13, 14–15 (2022).

[56] *See* Tilli Ripp & Jan Philipp Röer, *Systematic Review on the Association of COVID-19-Related Conspiracy Belief with Infection-Preventive Behavior and Vaccination Willingness*, 10 BMC PSYCH. 66, 66 (2022) ("Belief in COVID-19-related conspiracy narratives was negatively associated with vaccination willingness and infection-preventive behavior."); Sander van der Linden, Jon Roozenbeek & Josh Compton, *Inoculating Against Fake News About COVID-19*, 11 FRONTIERS PSYCH. 1, 2 (2020) (linking misinformation to distortion of risk perception and thus failure to adopt preventative measures). *But see* Marie Juanchich, Miroslav Sirota, Daniel Jolles & Lilith A. Whiley, *Are COVID-19 Conspiracies a Threat to Public Health? Psychological Characteristics and Health Protective Behaviours of Believers*, 51 EUR. J. OF SOC. PSYCH. 969, 969 (2021) ("Unexpectedly, COVID-19 conspiracy believers

stance, people who were exposed to COVID-19 misinformation had distorted views about the dangers posed by the virus, were less likely to comply with government public health guidance, had reduced inclination to wear masks, to adhere to other health-protective behavior, or to get vaccinated, and also had a tendency to encourage peers not to get vaccinated.[57] Some researchers have concluded that "health-related misinformation or disinformation can lead to more infections, deaths, disruption, and disorganization of the effort."[58] Though it's difficult to establish a causal connection between online PHM and these adverse effects, many recent studies imply that online misinformation has negative public health effects.[59]

---

adhered to basic health guidelines and advanced health protective measures as strictly as nonbelievers.").

[57] *See* Jon Roozenbeek, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra L.J. Freeman, Gabriel Recchia, Anne Marthe van der Bles & Sander van der Linden, *Susceptibility to Misinformation About COVID-19 Around the World*, 7 ROYAL SOC'Y OPEN SCI. 201199, 201199 ("[I]ncreased susceptibility to misinformation negatively affects people's self-reported compliance with public health guidance about COVID-19, as well as people's willingness to get vaccinated against the virus and to recommend the vaccine to vulnerable friends and family."); *see also* Daniel Freeman, Felicity Waite, Laina Rosebrock, Ariane Petit, Chiara Causier, Anna East, Lucy Jenner, Ashley-Louise Teale, Lydia Carr, Sophie Mulhall, Emily Bold & Sinéad Lambe, *Coronavirus Conspiracy Beliefs, Mistrust, and Compliance with Government Guidelines in England*, 52 PSYCH. MED. 251, 252 (2020) (finding people who are susceptible to COVID-19 misinformation are less likely to comply with government public health guidance, such as guidance on social contact); Marios Constantinou, Antonios Kagialis & Maria Karekla, *COVID-19 Scientific Facts vs. Conspiracy Theories: Is Science Failing to Pass its Message?*, 18 INT'L J. ENV'T. RSCH. & PUB. HEALTH 6343, 6343 (2021) ("Stronger conspiracy theory beliefs predicted science mistrust and unwillingness to adhere to public health measures."); Mehdi Mourali & Carly Drake, *The Challenge of Debunking Health Misinformation in Dynamic Social Media Conversations: Online Randomized Study of Public Masking During COVID-19*, 24 J. MED. INTERNET RSCH. E34831, 11 (2022) ("We found that exposure to misinformation has a negative impact on attitudes and intentions toward masking.").

[58] TARA KIRK SELL, DIVYA HOSANGADI, ELIZABETH SMITH, MARC TROTOCHAUD, PRARTHANA VASUDEVAN, GIGI KWIK GRONVALL, YONAIRA RIVERA, JEANNETTE SUTTON, ALEX RUIZ & ANITA CICERO, JOHNS HOPKINS BLOOMBERG SCH. PUB. HEALTH CTR. FOR HEALTH SEC., NATIONAL PRIORITIES TO COMBAT MISINFORMATION AND DISINFORMATION FOR COVID-19 AND FUTURE PUBLIC HEALTH THREATS: A CALL FOR A NATIONAL STRATEGY iii (2021).

[59] Ingjerd Skafle, Anders Nordahl-Hansen, Daniel S Quintana, Rolf Wynn & Elia Gabarron, *Misinformation About COVID-19 Vaccines on Social Media: Rapid Review*, 24 J. MED. INTERNET RSCH. E37367 (2022) (Eighteen of nineteen studies "implied that the misinformation spread on social media had a negative effect on vaccine hesitancy and uptake."). Some do explicitly point to the harms of online PHM. *See, e.g.*, Gallo & Cho, *supra* note 36, at 14–17; Francesco Pierri, Brea L. Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer & John Bryden, *Online Misinformation is Linked to Early COVID-19 Vaccination Hesitancy and Refusal*, 12 SCI. REPS. 5966, 5966 (2022) ("We find a negative relationship between misinformation and vaccination uptake rates. Online misinformation is also correlated with vaccine hesitancy rates taken from survey data."); Daniel Allington, Bobby Duffy, Simon Wessely, Nyana Dhavan & James Rubin, *Health-Protective Behaviour, Social Media Usage and Conspiracy Belief During the COVID-19 Public Health Emergency*, 51 PSYCH. MED. 1763 (finding "a negative relationship between COVID-19 conspiracy beliefs and COVID-19 health-protective behaviours, and a positive relationship between COVID-19 conspiracy beliefs and use of social media as a source of information about COVID-19").

Governments and leading global institutions have adopted these scientific findings, recognizing that misinformation during the COVID-19 pandemic was associated with negative opinions of vaccines and public-health advice, and a tendency not to follow recommended preventions and control behaviors.[60] For instance, the EU noted that PHM "can have severe consequences: it can lead people to ignore official health advice and engage in risky behaviour . . . . [It] directly endanger[s] lives and severely undermine[s] efforts to contain the pandemic."[61] In a similar vein, the Surgeon General of the United States concluded that: "[H]ealth misinformation is a serious threat to public health. It can cause confusion, sow mistrust, harm people's health, and undermine public health efforts. Limiting the spread of health misinformation is a moral and civic imperative."[62]

Nowhere is the connection between PHM and inferior public health outcomes more apparent, and more studied, than in the case of vaccine hesitancy. Needless to say, vaccines are an indispensable part of public health. They help save and improve lives around the world.[63] Noni E. MacDonald and the SAGE Working Group on Vaccine Hesitancy define vaccine hesitancy as "delay in acceptance or refusal of vaccination despite availability of vaccination services."[64] Vaccine hesitancy affects both the hesitant individual, who is not protected from the disease, and their entire community by undermining social endeavors to confer "herd immunity."[65] The World Health Organization ("WHO") identified vaccine hesitancy as one of its top-ten global health concerns.[66] PHM about the alleged harms or ineffectiveness of vaccines contributes to vaccine hesitancy and decreased intentions to get the vaccine.[67] These effects are heightened when groups that share such information use social media to organize offline action.[68]

One (in)famous case grimly illustrates the possible effects of PHM on close-knit communities. Following anti-vaccine advocates' visits to a Somali

---

[60] *See* Zarocostas, *supra* note 47; *see also* RAYNARD S. KINGTON, STACEY ARNESEN, WEN-YING SYLVIA CHOU, SUSAN J. CURRY, DAVID LAZER & ANTONIA M. VILLARRUEL, NAT'L ACAD. MED. PERSPS., IDENTIFYING CREDIBLE SOURCES OF HEALTH INFORMATION IN SOCIAL MEDIA: PRINCIPLES AND ATTRIBUTES 2 (2021).

[61] *EU Joint Communication*, *supra* note 47.

[62] SG REPORT, *supra* note 32, at 2.

[63] *See generally* ANA SANTOS RUTSCHMAN, VACCINES AS TECHNOLOGY: INNOVATION, BARRIERS, AND THE PUBLIC HEALTH (2022) (describing immense value of vaccine development and distribution).

[64] MacDonald, *supra* note 16, at 4163.

[65] Shimaa M. Saied, Eman M. Saied, Ibrahim Ali Kabbash & Sanaa Abd El-Fatah Abdo, *Vaccine Hesitancy: Beliefs and Barriers Associated with COVID-19 Vaccination Among Egyptian Medical Students*, 93 J. MED. VIROLOGY 4280, 4281 (2021).

[66] *See* MacDonald, *supra* note 16, at 4163.

[67] *See* Neha Puri, Eric A. Coomes, Hourmazd Haghbayan & Keith Gunaratne, *Social Media and Vaccine Hesitancy: New Updates for the Era of COVID-19 and Globalized Infectious Diseases*, 16 HUM. VACCINES & IMMUNOTHERAPEUTICS 2586, 2586 (2020); Massimiliano Mascherini & Sanna Nivakoski, *Social Media Use and Vaccine Hesitancy in the European Union*, 40 VACCINE 2215, 2215 (2022).

[68] *See* Steven Lloyd Wilson & Charles Wiysonge, *Social Media and Vaccine Hesitancy*, BMJ GLOB. HEALTH, Oct. 2020, at 5.

community in Minneapolis, that community's MMR vaccination levels dropped sharply, from ninety-two percent to forty-two percent in just one decade. Efforts by the health officials in the region were no match for the community's belief in PHM. As a result, wave after wave of terrible outbreaks of measles hit that community for years, causing illnesses and significant costs to public health.[69] To illustrate, in a 2017 large outbreak of measles in Minnesota, ninety-one percent of the infected were unvaccinated (eighty-one percent of infected were of Somali descent).[70] Out of seventy-five total cases, twenty-one children (all unvaccinated) were hospitalized with measles symptoms. In monetary terms, "[s]tate and key public health partners spent an estimated $2.3 million on response."[71] Emily Banerjee of the Minnesota department of health related the outbreak to PHM, noting:

> Misinformation about MMR vaccine continues to fuel vaccine hesitancy in Minnesota, the United States, and many other countries experiencing large measles outbreaks. . . . A collaborative global approach to promote and maintain high immunization rates, enhance public health infrastructure, and combat pervasive vaccine misinformation is crucial to stop measles from becoming endemic once again.[72]

III. Platforms' Efforts and Limitations

As the COVID-19 pandemic and infodemic raged, major platforms realized the harms of PHM and acted. They quickly and repeatedly updated their policies on misinformation and disinformation to address PHM, engaged in fact-checking and content moderation, promoted content they deemed reliable, and sanctioned users and groups that disseminated PHM.[73] Empirical studies suggest that platforms' efforts were fruitful—making PHM less common since the COVID-19 pandemic compared to before the pandemic.[74] Recent legal scholarship comprehensively surveyed these efforts.[75]

---

[69] O'Connor & Weatherall, *supra* note 17, at 142–43; Reiss & Diamond, *supra* note 17, at 544–53.

[70] Emily Banerjee, Jayne Griffith, Cynthia Kenyon, Ben Christianson, Anna Strain, K. Martin, Melissa McMahon, Erica Bagstad, E. Laine, Kristin Hardy, Genny Grilli, Jacy Walters, Denise Dunn, Margo Roddy & Kris Ehresmann, *Containing a Measles Outbreak in Minnesota, 2017: Methods and Challenges*, 140 Persps. Pub. Health 162, 162 (2020).

[71] *Id.*

[72] *Id.* at 170.

[73] *See, e.g.*, Kelley Cotter, Julia R. DeCook & Shaheen Kanthawala, *Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth*, 8 Soc. Media + Soc'y 1, 4 (2022).

[74] *See generally* David A. Broniatowski, Daniel Kerchner, Fouzia Farooq, Xiaolei Huang, Amelia M. Jamison, Mark Dredze, Sandra Crouse Quinn & John W. Ayers, *Twitter and Facebook Posts About COVID-19 Are Less Likely to Spread Misinformation Compared to Other Health Topics*, PLoS ONE, Jan. 12, 2022, at 1.

[75] *See, e.g.*, Nandita Krishnan, Jiayan Gu, Rebekah Tromble & Lorien C. Abroms, *Research Note: Examining How Various Social Media Platforms Have Responded to COVID-19 Misinformation*, 2 Harv. Kennedy Sch. Misinformation Rev. (2021), https://misinforeview.hks.harvard.edu/wp-content/uploads/2021/12/

This section will quickly survey the main efforts taken by major platforms and focus on discussing them. Briefly, we commend platforms' reactions to PHM during COVID-19 and explain that platforms acted as well as one could hope for. However, these worthy intentions and actions fell short, primarily due to a series of structural limitations platforms face. In this Section, then, we explain why government involvement in confronting online PHM is needed, even when platforms are at their best.

A. Platforms Actions Against Online PHM During COVID-19

Facebook has been quick to respond to the emerging infodemic regarding COVID-19, setting the support of COVID-19 vaccine rollouts as a top priority.[76] Starting in March 2020, the company cooperated with the WHO and local health agencies to propagate reliable information and confront PHM.[77] The platform opened and cultivated a COVID-19 information center, compiled extensive lists of prohibited PHM claims, reportedly removed over 20 million posts and 3,000 accounts and groups that spread PHM, posted warning labels on posts that included PHM and directed users to accurate information about it, and notified users that interacted with PHM.[78] But Facebook's response had limits. Its content moderation tools neglected comments to posts, through which PHM polluted many reliable posts about vaccine information.[79] And, until October 2020, the platform allowed PHM about COVID-19 vaccines in its ads.[80] By July 2022, as the risks posed by

---

krishnan_social_media_covid_19_20211215.pdf [https://perma.cc/Q3NP-YYMX]; Nunziato, *supra* note 6, at 37–51; Ana Santos Rutschman, *Social Media Self-Regulation and the Rise of Vaccine Misinformation*, 4 J.L. & INNOVATION 25, 43–57 (2021) [hereinafter Rutschman, *Self-Regulation*].

[76] *See* Sam Schechner, Glazer Jeff Horwitz & Emily Glazer, *How Facebook Hobbled Mark Zuckerberg's Bid to Get America Vaccinated*, WALL ST. J. (Sept. 17, 2021), https://www.wsj.com/articles/facebook-mark-zuckerberg-vaccinated-11631880296 [https://perma.cc/QB3X-RTYQ].

[77] Mark Zuckerberg, FACEBOOK (Mar. 3, 2020), https://www.facebook.com/4/posts/i-wanted-to-share-an-update-on-the-steps-were-taking-to-respond-to-the-coronavir/10111615249124441/ [https://perma.cc/W8FH-CVXJ].

[78] *See* Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, META (Apr. 16, 2020), https://about.fb.com/news/2020/04/covid-19-misinfo-update/ [https://perma.cc/TMM9-N938]; Naomi Nix & Kurt Wagner, *Facebook Removed 20 Million Pieces of COVID-19 Misinformation*, BLOOMBERG (July 1, 2022), https://www.bloomberg.com/news/articles/2021-08-18/facebook-removed-20-million-pieces-of-covid-19-misinformation [https://perma.cc/VUQ2-TX6R]; *COVID-19 Information Center*, META, https://about.meta.com/covid-19-information-center [https://perma.cc/U7KC-2HTY] .

[79] *See* Esther Chan, Lucinda Beaman & Stevie Zhang, *Vaccine Misinformation in Facebook Comment Sections: A Case Study*, FIRST DRAFT (May 6, 2021), https://firstdraftnews.org:443/articles/vaccine-misinformation-in-facebook-comment-sections-a-case-study/ [https://perma.cc/7WKQ-H88P].

[80] Caroline Haskins, *Facebook Is Running Anti-Vax Ads, Despite Its Ban on Vaccine Misinformation*, BUZZFEED NEWS (Jan. 8, 2020), https://www.buzzfeednews.com/article/carolinehaskins1/facebook-running-anti-vax-ads-despite-ban-anti [https://perma.cc/E42X-M7L4]; *Vaccine Discouragement*, FACEBOOK, https://www.facebook.com/policies/ads/

COVID-19 started to wane, Facebook considered rolling back some of its policies to confront PHM.[81]

YouTube and Twitter had similarly mixed success with PHM during the pandemic. Twitter also contacted the WHO and governmental sources to identify reliable public health information.[82] It developed policies to label or remove harmful or misleading content, and directed users to reliable content instead.[83] It also provided specific NGOs and non-profits pro-bono advertising to help them spread reliable information in many countries.[84] Between January 2020 and September 2022, Twitter suspended over 11,000 accounts and removed over 97,000 pieces of contents based on its COVID-19 guidance.[85] Twitter, however, has less friction than other platforms for sharing PHM, making misinformation amplification easier on the platform.[86] Beginning November 23, 2022, however, Twitter stopped enforcing its COVID-19 misleading information policy.[87]

YouTube also made considerable efforts to confront PHM. The platform adopted a "COVID-19 medical misinformation policy."[88] It prohibits content that "spreads medical misinformation that contradicts local health authorities[ ] and the WHO[ ]," and features public health information from private organizations.[89] Despite these efforts, YouTube has been a host to an abundance of PHM. The virality of the video *Plandemic*, which claimed a cabal of powerful elites are behind the pandemic, illustrates this point.[90]

---

prohibited_content/vaccine_discouragement [https://perma.cc/FF7M-G5QA] (prohibiting advertisements that discourage vaccination).

[81] *See* Nick Clegg, *Meta Asks Oversight Board to Advise on COVID-19 Misinformation Policies*, META (July 26, 2022), https://about.fb.com/news/2022/07/oversight-board-advise-COVID-19-misinformation-measures/?utm_source=substack&utm_medium=email [https://perma.cc/5QSG-2UPV].

[82] Twitter Public Policy, *Stepping Up Our Work to Protect the Public Conversation Around COVID-19*, TWITTER BLOG (Mar. 4, 2020), https://blog.twitter.com/en_us/topics/company/2020/stepping-up-our-work-to-protect-the-public-conversation-around-covid-19 [https://perma.cc/EEX4-B672].

[83] Yoel Roth & Nick Pickles, *Updating Our Approach to Misleading Information*, TWITTER BLOG (May 11, 2020), https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information [https://perma.cc/NCA8-LZGW].

[84] Twitter Safety, *Updates to Our Work on COVID-19 Vaccine Misinformation*, TWITTER (Mar. 1, 2021), https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation [https://perma.cc/KD6U-PPU8].

[85] *COVID-19 Misinformation*, TWITTER TRANSPARENCY, https://transparency.twitter.com/en/reports/covid19.html#item0:2021-jan-jun: [https://perma.cc/MB2B-RB3X].

[86] *See Misinformation Amplification Analysis and Tracking Dashboard*, INTEGRITY INST. (Oct. 13, 2022), https://integrityinstitute.org/our-ideas/hear-from-our-fellows/misinformation-amplification-tracking-dashboard [https://perma.cc/5LXE-RTND].

[87] *See* Donie O'Sullivan, *Twitter Is No Longer Enforcing Its COVID Misinformation Policy*, CNN BUSINESS (Nov. 29, 2022), https://www.cnn.com/2022/11/29/tech/twitter-covid-misinformation-policy/index.html [https://perma.cc/WG9M-JYCH].

[88] *Covid-19 medical misinformation policy*, YOUTUBE HELP, https://support.google.com/youtube/answer/9891785?hl=EN [https://perma.cc/QQ22-WX45] (last visited Feb. 28, 2022).

[89] *Id.*

[90] Grant Currin, *YouTube's Plan to Showcase Credible Health Information Is Flawed, Experts Warn*, SCI. AM. (Aug. 27, 2021), https://www.scientificamerican.com/article/youtubes-

B. Why Private Ordering Is Insufficient

Indeed, platforms made commendable efforts to regulate PHM in the midst of the COVID-19 pandemic.[91] In this respect, COVID-19 PHM is a case study in what platforms are willing to do to confront misinformation. However, this case study also highlights a few important limitations and shortcomings of relying only on platforms for confronting PHM.

First, all platforms faced a structural problem—forceful action against PHM would undermine their business interests.[92] Hindering PHM about COVID-19 would cut against their core business model of promoting content that optimizes user engagement.[93] This in turn steers users to more extreme or radical versions of the content they find interesting.[94] Platforms that rely on advertising revenue—as all major platforms do—are likely to engage in content moderation, but with lax community standards, in order to retain a larger group of consumers.[95] As Imran Ahmed, the CEO of the Center for Countering Digital Hate put it: "Why would you not remove comments? Because engagement is the only thing that matters. It drives attention and attention equals eyeballs and eyeballs equal ad revenue."[96] Put simply, platforms' business interests are often structurally misaligned with the public interests regarding confronting PHM.

Second, platforms were selectively transparent about their efforts to confront PHM. For instance, while platforms were keen on reporting their actions against PHM,[97] they were often reluctant to share the amount of PHM on their platforms.[98] This is the denominator problem: without the lat-

---

plan-to-showcase-credible-health-information-is-flawed-experts-warn/ [https://perma.cc/EK3J-PDVX].

[91] Krishnan et al., *supra* note 75, at 1.

[92] *See infra* note 331 (discussing Facebook's limitations in countering PHM during the COVID-19 pandemic).

[93] *Cf.* Laura Edelson, Minh-Kha Nguyen, Ian Goldstein, Oana Goga, Tobias Lauinger & Damon McCoy, *Understanding Engagement with (Mis)Information News Sources on Facebook*, MEDIUM (Sept. 14, 2021), https://medium.com/cybersecurity-for-democracy/understanding-engagement-with-mis-information-news-sources-on-facebook-8d39bca38978 [https://perma.cc/8EUH-U8BY] (measuring more user engagement with posts from misinformation news sources); Mathew Ingram, *YouTube Has Done Too Little, Too Late to Fight Misinformation*, COLUM. JOURNALISM REV. (Apr. 4, 2019), https://www.cjr.org/the_media_today/youtube-misinformation.php [https://perma.cc/S55V-KBMH] (arguing YouTube CEO prioritized increasing engagement over curbing misinformation).

[94] Katherine J. Wu, *Radical Ideas Spread Through Social Media. Are the Algorithms to Blame?* NOVA (Mar. 28, 2019), https://www.pbs.org/wgbh/nova/article/radical-ideas-social-media-algorithms/ [https://perma.cc/C6ZG-6KX8].

[95] *See* Yi Liu, Pinar Yildirim & Z. John Zhang, *Implications of Revenue Models and Technology for Content Moderation Strategies*, 41 MKTG. SCI. 831, 833 (2022).

[96] David Klepper & Amanda Seitz, *Facebook Froze as Anti-Vaccine Comments Swarmed Users*, AP NEWS (Oct. 26, 2021), https://apnews.com/article/the-facebook-papers-covid-vaccine-misinformation-c8bbc569be7cc2ca583dadb4236a0613 [https://perma.cc/48U5-8G4H].

[97] *See supra* notes 78, 85.

[98] Gerrit De Vynck, Cat Zakrzewski & Cristiano Lima, *Facebook Told the White House to Focus on the 'Facts' About Vaccine Misinformation. Internal Documents Show It Wasn't*

ter information, the former is merely a number that tells us very little about the policies and their effects. Platforms' disinclination to admit PHM is being disseminated uncontrolled under their watch is understandable. Thus, yet again, platforms' business interests seem misaligned with public interests, this time on information sharing about PHM.

Third, regulating PHM involves epistemic questions that should not be resolved by platforms. As a practical matter, simply 'relying on science' is not enough: someone has to read and make sense of the scientific sources about the given issue, and provide an authoritative answer to practical questions.[99] Deciding which actors we can and should trust to make these calls is a complicated question in political epistemology.[100] Governmental institutions like the C.D.C. regularly fulfill this role, acting as indispensable sources of reliable information about public health. But platforms are not bound to defer to those agencies. They can easily ignore governmental views and opt for other sources like a group of randomly selected users. Twitter's quick policy changes under its new ownership,[101] and TikTok's rise to prominence despite controversial policies and allegiances,[102] make this point evident. Acknowledging this epistemic problem raises a dilemma. Should we trust specialized government agencies or whoever the platform decides to trust? We find no epistemic or political reasons to empower private companies to solve those epistemic challenges. While we recognize the shortcomings of existing government institutions, there are good reasons to trust those institutions. Thus, we choose government, for reasons that will become clear shortly.

Fourth, regulating online speech, including PHM, in a democratic society raises complicated normative tradeoffs and considerations. Those questions, we argue, should not be left to the private companies running online platforms. One question involves equal treatment of speakers. At their finest, social media platforms allow socially marginalized individuals and minority groups to voice their concerns, gather support (and suffer criticism). Plat-

---

Sharing Key Data, WASH. POST (Oct. 28, 2021), https://www.washingtonpost.com/technology/2021/10/28/facebook-covid-misinformation/ [https://perma.cc/7CQB-2RVS].

[99] *See generally* Thomas Grundmann, *Experts: What Are They and How Can Laypeople Identify Them?*, *in* OXFORD HANDBOOK SOC. EPISTEMOLOGY 12 (Jennifer Lackey & Aidan McGlynn eds., forthcoming 2023).

[100] *See supra* note 24 and accompanying text. On political epistemology, see generally THE ROUTLEDGE HANDBOOK OF POLITICAL EPISTEMOLOGY (Michael Hannon & Jeroen de Ridder eds., 2021).

[101] *See, e.g.*, Mark MacCarthy, *How Elon Musk Might Shift Twitter Content Moderation*, BROOKINGS (Apr. 29, 2022), https://www.brookings.edu/blog/techtank/2022/04/29/how-elon-musk-might-shift-twitter-content-moderation/ [https://perma.cc/WJ4F-CPEN]; Jack Brewster, Macrina Wang & Valerie Pavilonis, *Twitter Misinformation Superspreaders See Huge Spike in Engagement Post-Acquisition by Elon Musk*, NEWSGUARD (Nov. 11, 2022), https://www.newsguardtech.com/special-reports/twitter-misinformation-superspreaders-see-huge-spike-in-engagement-post-acquisition-by-elon-musk/ [https://perma.cc/7JB7-YZU4].

[102] *The All-Conquering Quaver*, ECONOMIST (July 9, 2022), https://www.economist.com/interactive/briefing/2022/07/09/the-all-conquering-quaver [https://perma.cc/8EUH-U8BY] (discussing the rise of TikTok and its ties to the Chinese government).

forms magnify voices that might otherwise be too silent. Arguably, this reasoning should protect both #MeToo advocates' and vaccine hesitants' online speech. Any rule distinguishing between the two groups or supporting only one's speech interests on platforms would be difficult to administer, especially if both topics are politically contested and if platforms want to preserve content-neutrality. Another normative problem is inherent to PHM. No one thinks that all falsities should be prohibited online. Societies should tolerate some level of false speech in the public realm, including with regards to PHM.[103] But how much? Determining the appropriate level of tolerance to false speech, and whether to sanction violators (if at all), are contested normative questions. And even after we craft perfect policies, other questions arise. Since no enforcement tool would be perfect in applying such policies, mistakes about the application of those policies—allowing speech that should be removed, or suppressing speech that should be allowed—are inevitable. So, an additional question arises about the allocation of mistakes: should platforms err on the side of allowing content or restricting it?

The questions posed in the previous paragraph are hardly novel. They are a new iteration of classic normative questions about speech regulation. They linger because they are both persistent in human interactions and not easily solved. In a democracy, we resolve these normative questions through the political process and representative decision-making by elected officials.[104]

Leaving these complicated normative questions about regulation of PHM solely in the hands of online platforms is problematic. To begin, platforms are not set up to make justified political decisions. They are unaccountable private companies, driven primarily by economic incentives and the desire to make profits. Additionally, individuals' and polities' ability to influence and participate in crafting platforms' policies is limited at best. Needless to say, participation goes a long way in political justification.[105] Allowing relevant parties to participate in a decision (that is, to influence, comment, make arguments, provide information, and become informed), renders that decision justified to the participating actors. And vice versa, when polity members are barred from participation, they may always object to the decision and protest its binding force on them.[106] Of course, private

---

[103] Reasons to do so include valuing human expression even when it's false, not knowing the truth, or avoiding inadvertently chilling valuable speech. *See* United States v. Alvarez, 567 U.S. 709, 723 (2012); *infra* Section IV.F.

[104] *See generally* JEREMY WALDRON, LAW AND DISAGREEMENT 88–118 (1999) (exploring the processes and voting involved in representative legislatures).

[105] On the importance of participation in politics and democracy, see, e.g., *id.* at 232–54; Lawrence B. Solum, *Procedural Justice*, 78 CAL. L. REV. 181, 279–81 (2004).

[106] *See generally* Daniel Markovits, *Adversary Advocacy and the Authority of Adjudication*, 75 FORDHAM L. REV. 1367, 1374–78, 1384–86 (2006); Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 354 (1978); THOMAS HOBBES, LEVIATHAN (Prometheus ed., 1988) (1651); HART & SACKS, THE LEGAL PROCESS 640–47 (1951); Solum, *Procedural Justice*, *supra* note 105, at 279–84.

platforms should not be faulted for not facilitating political participation (or any other political good). It is simply not in their nature and purpose. However, absent such justifying characteristics, why should platforms be empowered to make important normative decisions about online speech, decisions which affect the entire polity?

We don't think they should. We find arguments for favoring platforms as the governors of online speech unconvincing. Those like us, who feel discomfort with platforms' dominance as uncontested "new governors"[107] should welcome governments' role in regulating online PHM.

Let's conclude. Online PHM poses considerable harms to individuals and societies, harms that intensify in times of public health hazards such as epidemics or pandemics.[108] There are good normative reasons to favor government's involvement in the efforts to confront online PHM: governments are the primary tools through which societies confront social challenges, societies have mechanisms to influence what governments do and to participate in the decision-making process, governments have professional knowledge about public health, and governments' actions are checked and balanced by developed institutional frameworks. None of these reasons apply to platforms. Indeed, we see no compelling normative reasons to completely privatize the social efforts to confront online PHM. This argument holds when platforms act in good faith and with competence, as big platforms mostly did during the COVID-19 pandemic. It is heightened when we consider that nothing guarantees platforms won't change their views, leadership, or policies tomorrow.[109] Governmental involvement is a particularly viable option when the desired policies might undermine platforms' business interests.[110]

In confronting online PHM, we favor governments' involvement over reliance solely on unregulated private platforms. As we shall see in the next Part, government has many ways to influence how platforms govern PHM. Unfortunately, many of the existing approaches fall short.

IV. Liability for COVID-19 Misinformation: The Limits of Existing Approaches

So far, we have explained that online PHM causes considerable harms to individuals and societies, explained the limitations of platforms' responses, and argued that platforms should not confront this problem exclusively. We now turn to a brief review of existing legal tools for confronting

---

[107] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 1599 (2018).

[108] *See supra* Section II.B.

[109] *See supra* notes 101–02.

[110] *See supra* Section III.B.

online PHM, finding serious flaws in each of these approaches. This analysis illustrates the importance of the new approaches we will suggest in Part V.

*****

Are individuals who create and share COVID-19 misinformation legally responsible for any harm they cause to others? And what about the platforms that help amplify their message? The short answer is: only to a very limited extent.

First, public health law empowers the state to take strict measures in order to promote public health, but its tools for addressing online PHM are limited. Second, consumer protection laws empower the Federal Trade Commission ("FTC"), the Food and Drug Administration ("FDA"), and state Attorneys General to police false or fraudulent PHM claims by companies. But those focus only on commercial transactions, which constitute a minor share of online PHM. Third, medical licensing authorities can theoretically help, but in practice they seldom act against doctors that disseminate PHM. And their mandate is limited only to those in the medical profession. Fourth, the tort regime offers redress to those injured by advice or treatment premised on PHM, but only if health professionals offer them. If PHM harms a person, they can bring a negligent misrepresentation claim, but they face an uphill battle.[111] Fifth, the First Amendment prohibits most attempts to outright regulate false speech.[112] The millions of ordinary citizens who use social media to share their views on the various aspects of the COVID-19 pandemic enjoy First Amendment protection even when they disseminate PHM. Sixth, Section 230 of the Communications Decency Act broadly protects social media platforms from liability for publishing or removing user-generated content.[113] Hence, even if a plaintiff's legal claim for harms caused by PHM miraculously prevails, platforms are protected.

A. Public Health Law

Public health law generally refers to laws and regulations that enhance public health, reduce health hazards and risk factors, and assure that people and populations are healthy.[114] Public health law focuses on actions and health benefits of the polity as a whole.[115] Achieving public health requires

---

[111] *See generally* Reiss & Diamond, *Measles and Misrepresentation in Minnesota*, *supra* note 17.

[112] *See* United States v. Alvarez, 567 U.S. 709, 715 (2012).

[113] 47 U.S.C. § 230 (2018).

[114] *See generally* Gostin & Wiley, *supra* note 54, at 12–16; Hodge, *supra* note 53, at 12–17.

[115] Public health itself can be defined as "the science and art of preventing disease, prolonging life, and promoting health through the organized efforts and informed choices of society, organizations, public and private communities, and individuals." *Introduction to*

social ordering and political decision-making.[116] Public health law empowers the government to take various actions that undermine individuals' interests and rights to "safeguard the public health."[117] During the COVID-19 pandemic, courts reiterated the public interest in confronting public health hazards, holding that "few interests are more compelling than protecting public health against a deadly virus."[118] Public health law encapsulates various kinds of state action that attempt to promote public health. Those include: requiring vaccination;[119] imposing medical quarantine or isolation;[120] and declaring public health emergencies, which grants officials a myriad of additional powers.[121] Public health law also involves governments' issuing guidelines that regulate others' speech. These guidelines include specific labeling requirements for manufacturers and sellers, restrictions on misleading or false advertising, and official letters requiring actors to stop selling dangerous products.[122]

Two points are worth noting. First, government actors regularly communicate public health messages—promoting safer behaviors, healthy eat-

---

Public Health, CDC (quoting C.E.A. Winslow), https://www.cdc.gov/training/publichealth101/public-health.html [https://perma.cc/5FTP-MHGL].

[116] *See* GOSTIN & WILEY, *supra* note 54, at 5–10. On the need for social ordering in order to achieve shared social goals and address collective action problems, *see generally* HART & SACKS, THE LEGAL PROCESS, *supra* note 106.

[117] Jacobson v. Massachusetts, 197 U.S. 11, 25 (1905); *see generally* GOSTIN & WILEY, *supra* note 54.

[118] Does 1-6 v. Mills, 16 F.4th 20, 32 (1st Cir. 2021), *cert. denied sub nom.* Does 1-3 v. Mills, 142 S. Ct. 1112 (2022) (mem.).

[119] *Jacobson*, 197 U.S. at 34. For recent applications, see, e.g., Biden v. Missouri, 142 S. Ct. 647, 650 (2022) (finding Secretary of Health and Human Services can require staff at Medicare and Medicaid participating facilities to get vaccinated, unless exempt for medical or religious reasons); Workman v. Mingo Cty. Bd. of Educ., 419 F. App'x 348, 353–54 (4th Cir. 2011) ("[F]ollowing the reasoning of *Jacobson* and *Prince*, we conclude that the West Virginia statute requiring vaccinations as a condition of admission to school does not unconstitutionally infringe Workman's right to free exercise.").

[120] *See* HODGE, *supra* note 53, at 147–62; *see also* Marie Sutton, *Forced Quarantine & Isolation: Does the Law Adequately Balance Individual Rights and Societal Protection?*, 39 U. LA VERNE L. REV. 98, 104–15 (2017) (discussing the regulatory framework governing the power to impose quarantine under public health law); *Jacobson*, 197 U.S. at 29. For cases upholding officials' stay-at-home orders, see, e.g., Williams v. Trump, 495 F. Supp. 3d 673, 682 (N.D. Ill. 2020), *aff'd sub nom.* Williams v. Pritzker, No. 20-3231, 2021 WL 4955683 (7th Cir. Oct. 26, 2021); Lawrence v. Polis, 505 F. Supp. 3d 1136, 1139 (D. Colo. 2020); Hartman v. Acton, 499 F. Supp. 3d 523, 528 (S.D. Ohio 2020).

[121] *See, e.g.*, James G. Hodge, Jr., Sarah A. Wetter & Erica N. White, *Legal Crises in Public Health*, 47 J.L. MED. & ETHICS 778, 778 (2019); James G. Hodge, Jennifer L. Piatt, Hanna N. Reinke & Emily Carey, *COVID's Constitutional Conundrum: Assessing Individual Rights in Public Health Emergencies*, 88 TENN. L. REV. 837 (2021) (discussing the various declarations, the public health measures taken, and many of the legal challenges to those declarations). For limitations of public health powers, especially when they hinder religious liberty, see, e.g., Calvary Chapel Dayton Valley v. Sisolak, 140 S. Ct. 2603, 2608 (2020) (Alito, J., dissenting); Roman Cath. Diocese of Brooklyn v. Cuomo, 141 S. Ct. 63, 68 (2020) (enjoining limitations on gatherings in religious sites based on First Amendment free exercise claims); Tandon v. Newsom, 141 S. Ct. 1294, 1296 (2021) (enjoining orders limiting in-home gathering during COVID-19 as burdening free exercise clause).

[122] *See, e.g.*, HODGE, *supra* note 53, at 308–27; *infra* notes 164–65 and accompanying text.

ing, physical activities, vaccinations, and so on.[123] Public health law recognizes that government can determine and propagate information about health, even if the information turns out to be false in hindsight.[124] Put simply, the possibility of getting the facts about public health wrong does not prohibit the government from adopting policies to confront contagious diseases.[125] Second, public health law, as its name suggests, focuses on the public's needs and takes a "population-based approach."[126] This approach affects possible speech regulation. Public health law's emphasis on public interests may serve as a counter-measure to individuals' (or companies') freedom of speech. This balancing act, for instance, justified bans on advertising tobacco products near schools.[127] Interestingly, this approach is harmonious with recent calls to favor audience's interests within the free speech discussion, especially on social media where speech is "cheap" and abundant.[128]

The previous paragraphs sound optimistic. It appears as if public health law has it all covered. Unfortunately, this is not the case. Public health law is primarily a legal framework. It is an umbrella term that incorporates many legal doctrines with regard to our topic. As such, public health law is helpful in creating emergency powers, requiring quarantine or imposing sanctions for not getting vaccinated (though those were recently criticized by the Supreme Court).[129] But it cannot do much to affect online PHM. Public health law provides ample justification and support for counter-PHM measures. But those justifications often fail in light of specific legal doctrines or the all-encompassing protections of the First Amendment. So, in practice, other than disseminating information, public health law's ability to stop the spread of PHM is miniscule. The following pages will concretize this argument.

B. Government Speech

> "'Government speech' refers to a wide range of phenomena in which, rather than regulating private speakers' messages, the government controls or supports a particular message using any of a panoply of carrots (such as funding or special access to govern-

---

[123] *See, e.g.*, Gostin & Wiley, *supra* note 54, at 141–43, 435–76; Parmet & Smith, *supra* note 49, at 373–90; Hodge, *supra* note 53, at 300–27.

[124] *See Jacobson*, 197 U.S. at 30; S. Bay United Pentecostal Church v. Newsom, 140 S. Ct. 1613, 1613–14 (2020) (Roberts, C.J., concurring); S. Bay United Pentecostal Church v. Newsom, 141 S. Ct. 716, 716–17 (2021); Andino v. Middleton, 141 S. Ct. 9, 10 (2020).

[125] *Jacobson*, 197 U.S. at 35.

[126] Parmet & Smith, *supra* note 49, at 436–40.

[127] *See, e.g.*, Lorillard Tobacco Co. v. Reilly, 533 U.S. 525, 565–66 (2001) (finding restrictions on tobacco advertising within one thousand feet of schools and playgrounds unconstitutional).

[128] *See generally* Richard L. Hasen, Cheap Speech: How Disinformation Poisons Our Politics—and How to Cure it (2022).

[129] *See, e.g.*, *Jacobson*, 197 U.S. at 24. *See generally* Hodge, *supra* note 53.

> ment property) or sticks (such as denial of funding or exclusion from government property)."[130]

Government speech is protected from most First Amendment claims.[131] That means, among other things, that the government does not have to remain neutral when it speaks.[132] The rationale for protecting government speech is practical and difficult to deny: the government must be able to convey its messages and to express the doctrines it holds in order to govern.[133] Government speech also empowers the government to discover and propagate to the public reliable and useful information, uncontaminated by business interests.[134] It also allows the government to inform and explain its actions—which in turn helps drive people to action without use of force.[135]

Government speech is essential for public health. For instance, getting each individual to wear their mask above the nose and mouth is almost impossible without it.[136] Governments often engage in health communication campaigns, explaining and convincing the public in order to promote public health goals, and alerting the public about health risks.[137] The aim of these educational efforts is to promote preventive care such as healthier diets, getting vaccinated, stopping smoking, etc.[138] As hinted above, government speech about public health issues can drive individuals and communities to

---

[130] B. Jessie Hill, *Introduction: Government Speech*, 61 CASE W. RSRV. L. REV. 1081, 1082 (2010).

[131] *See* Frederick Schauer, *Is Government Speech a Problem?*, 35 STAN. L. REV. 373, 383–86 (1983) (noting it is unclear why government speech should even be considered a First Amendment problem).

[132] Hill, *supra* note 130, at 1083 ("No government could do its job, after all, if it had to provide a podium for opposing views whenever it expressed its own views on matters like foreign policy or public health.").

[133] Robert C. Post, *Between Governance and Management: The History and Theory of the Public Forum*, 34 UCLA L. REV. 1713, 1825–26 (1987) ("Government organizations would grind to a halt were the Court seriously to prohibit viewpoint discrimination in the internal management of speech.").

[134] *See, e.g.*, CDC ENTERPRISE SOCIAL MEDIA POLICY (Jan. 8, 2015), https://www.cdc.gov/maso/policy/SocialMediaPolicy508.pdf [https://perma.cc/H6XB-ANUF]; CDC, FACEBOOK (Apr. 9, 2022), https://www.facebook.com/cdc/posts/354133043414807 [https://perma.cc/XC87-B7WW].

[135] *See, e.g.*, Helen Norton, *Constraining Public Employee Speech: Government's Control of Its Workers' Speech To Protect Its Own Expression*, 59 DUKE L.J. 1, 21–22 (2009); THOMAS I. EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 697–98 (1970).

[136] For example, airline companies have faced enormous problems with enforcing mask mandates on planes. *See* Letter from the Union of Southwest Airlines Flight Attendants, to Joseph R. Biden, President of the United States (Mar. 22, 2022), https://twu556.org/wp-content/uploads/formidable/54/MMLocal556.pdf [perma.cc/7HL4-5AK7] ("Serving onboard during these contentious times and enforcing mask compliance is one of the most difficult jobs we have ever faced as flight attendants . . . . The number of physical and verbal assaults in our workplace has increased dramatically, many of which are related to mask compliance.").

[137] For instance, during the COVID-19 pandemic the CDC issued guidance and toolkits with information for pregnant people and for children, as well as data trackers. *See COVID-19 Toolkits*, CDC (Apr. 11, 2022), https://www.cdc.gov/coronavirus/2019-ncov/communication/toolkits/index.html [https://perma.cc/8A6V-J5E7].

[138] *See* GOSTIN & WILEY, *supra* note 54, at 15–16, 141–42; HODGE, *supra* note 53, at 187–89.

make choices that are more conducive to public health.[139] It does so by increasing individuals' knowledge about risk conditions and increasing awareness and availability of valuable public health information.[140]

Government speech is also a valuable method in confronting online PHM.[141] The CDC has long been using social media as "a strategic communications tool," which allows "increasing the dissemination and potential impact of CDC's science [and] . . . [e]nhancing health communication efforts."[142] During the COVID-19 pandemic, the CDC and FDA used various platforms as well as their own websites to convey public health information and disseminate accurate and actionable advice to the public.[143] For instance, in response to an increasing trend (online and offline) to use ivermectin to treat COVID-19, the FDA published articles and tweets warning that this drug is not an effective treatment and is actually dangerous to humans.[144] In the United Kingdom, the government initiated a Rapid Response Unit that aimed to identify "false narratives" about COVID-19. Once identified, the unit responded by issuing "direct rebuttal on social media, working with platforms to remove harmful content and ensuring public health campaigns are promoted through reliable sources."[145]

Government speech can also be directed specifically to the platforms, in order to influence them to act against online PHM.[146] This includes publishing official open letters calling on platforms to promote public health goals, such as increasing confidence in COVID-19 vaccines.[147] Alternatively, gov-

---

[139] *See* Parmet & Smith, *supra* note 49, at 375–80.

[140] *See* HODGE, *supra* note 53, at 300–07.

[141] *See generally* Heidi J. Larson, *The Biggest Pandemic Risk? Viral Misinformation*, 562 NATURE 309 (2018); Heidi J. Larson, *Blocking Information on COVID-19 Can Fuel the Spread of Misinformation*, 580 NATURE 306 (2020).

[142] *See* CDC ENTERPRISE SOCIAL MEDIA POLICY, *supra* note 134.

[143] *See, e.g.*, CDC, FACEBOOK, *supra* note 134 ("Vaccinating children is the single best way to protect them from getting very sick with COVID-19. Learn more in this week's COVID Data Tracker Weekly Review: http://bit.ly/CDTweeklyreview.").

[144] *See, e.g.*, U.S. FDA (@US_FDA), TWITTER (Aug. 21, 2021), https://twitter.com/us_fda/status/1429050070243192839 [https://perma.cc/5FF6-3T82] ("You are not a horse. You are not a cow. Seriously, y'all. Stop it.").

[145] Press Release, UK Cabinet Office et. al., Government Cracks Down on Spread of False Coronavirus Information Online (Mar. 30, 2020), https://www.gov.uk/government/news/government-cracks-down-on-spread-of-false-coronavirus-information-online [https://perma.cc/2E3Z-XRJ4]; UK PARLIAMENT, MISINFORMATION IN THE COVID-19 INFODEMIC §§ 59–64, https://publications.parliament.uk/pa/cm5801/cmselect/cmcumeds/234/23406.htm#_idTextAnchor056 [https://perma.cc/WH8V-L6QU] (discussing the actions of the counter-disinformation unit with regards to COVID-19 misinformation).

[146] *Cf.* Derek E. Bambauer, *Orwell's Armchair*, 79 U. CHI. L. REV. 863, 891–99 (2012) (discussing governments' attempts to persuade platforms to act in some desired way).

[147] *See, e.g.*, Letter from Amy Klobuchar and Ben Ray Lujan, United States Senators, to Jack Dorsey, CEO of Twitter, and Mark Zuckerberg, CEO of Facebook (April 16, 2021), https://www.klobuchar.senate.gov/public/_cache/files/8/7/87e50146-a4cc-4ab1-9604-3190401bbec5/859B41CE812B8AC97F55D24EFEA0F834.4.16.21-letter-to-tech-ceos—vaccine-misinfo-final-.pdf [https://perma.cc/6Y9E-CPZF].

ernments can lobby platforms to confront PHM.[148] In a more adversarial manner, public officials can also summon platform executives to official hearings, publicly urging them to act against PHM in specific ways or requiring them to submit information about their actions.[149] Additionally, state-imposed sanctions can serve as a message to platforms (and other companies more generally) about how the government wants platforms to act on a specific issue.[150]

Of course, government public health speech is no panacea. Following internal reviews, the CDC found that it was often too slow to convey reliable, actionable information during the COVID-19 pandemic.[151] Government speech about public health issues might be false, and yet still be protected under the First Amendment.[152] As the COVID-19 pandemic made evident, elected officials, including the President, might abuse their powers to spread PHM.[153] These unfortunate efforts often lead to harmful, sometimes deadly, results for individuals and communities.[154] Additionally, professionally appointed public health officials can also spread PHM. This is particularly troubling because these actors speak from a position of dual authority—political and epistemic—and thus many regard them as reliable sources on public health issues.[155] We do not underestimate these risks. But those are

---

[148] *Cf.* Adam Satariano, *The World's First Ambassador to the Tech Industry*, N.Y. TIMES (Sept. 3, 2019), https://www.nytimes.com/2019/09/03/technology/denmark-tech-ambassador.html [https://perma.cc/D7GP-T4ZX] (discussing Denmark sending an ambassador to Facebook, among other tech companies).

[149] *See, e.g.*, *Disinformation Nation: Social Media's Role in Promoting Extremism and Misinformation Before the Subcomm. on Commc'n & Tech. of the H. Comm. on Energy & Commerce,* 117th Cong. (2021).

[150] *See, e.g.*, Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2327–29 (2014) (discussing governments' efforts to harness platforms and other companies to sanction WikiLeaks) [hereinafter Balkin, *New School*]; Bambauer, *Orwell's Armchair*, *supra* note 146, at 891–99; Derek E. Bambauer, *Against Jawboning*, 100 MINN. L. REV. 51, 65–83 (2015) (providing an overview of instances in which the United States government used its authority to persuade internet platforms to carry out its wishes); Jeffrey A. Sonnenfeld & Steven Tian, *Some of the Biggest Brands Are Leaving Russia. Others Just Can't Quit Putin. Here's a List.*, N.Y. TIMES (Apr. 7, 2022), https://www.nytimes.com/interactive/2022/04/07/opinion/companies-ukraine-boycott.html [https://perma.cc/248B-5BLP] (describing how many companies have reduced their footprints in Russia beyond what is legally required by government sanctions).

[151] *See e.g.*, Sharon LaFraniere & Noah Welland, *Walensky, Citing Botched Pandemic Response, Calls for C.D.C. Reorganization*, N.Y. TIMES (Aug. 17, 2022), https://www.nytimes.com/2022/08/17/us/politics/cdc-rochelle-walensky-covid.html [https://perma.cc/PVE4-Z42W].

[152] *See* Norton, *supra* note 135, at 23; Jonathan D. Varat, *Deception and the First Amendment: A Central, Complex, and Somewhat Curious Relationship*, 53 UCLA L. REV. 1107, 1132–33 (2006).

[153] *See, e.g.,* Claudia E. Haupt & Wendy E. Parmet, *Lethal Lies: Government Speech, Distorted Science, and the First Amendment*, 2022 U. ILL. L. REV. 1809, 1813–14 (2022) (discussing Trump's PHM about the nature and seriousness of the disease and about the efficacy and safety of certain treatments).

[154] *See, e.g.*, Jeffrey Kluger, *Accidental Poisonings Increased After President Trump's Disinfectant Comments*, TIME (May 12, 2020), https://time.com/5835244/accidental-poisonings-trump/ [https://perma.cc/A98B-M3BE].

[155] *See* Haupt & Parmet, *supra* note 153, at 1810–12.

ordinary risks of governmental power abuse. Democracy has rules and procedures in place to meet those risks, which it ordinarily relies on for abuses of power with greater potential harm (e.g., deploying the military or declaring emergencies). We see no reason to limit governmental public health speech in light of those risks.[156] So, government speech is, and will likely remain, a crucial tool for promoting public health goals, including confronting online PHM.

The more serious problem with relying on government speech to confront PHM, is that it simply does not seem to work. Online speech is mired by PHM and the individuals and NGOs that voluntarily try to confront it are vastly outnumbered[157] and often less prominent than the perpetrators.[158] Furthermore, the underlying theory for this approach seems to be that "[t]he remedy for speech that is false is speech that is true."[159] Hence, if only the government will publish accurate information, the problem of online PHM would be solved. But this marketplace of ideas assumption was always doubtful.[160] It seems even more questionable in an age of fast, amplifiable, and cheap online speech.[161] And more specifically, the marketplace of ideas metaphor seems particularly ill-suited for online PHM.

So, government speech about public health is necessary to confront the effects of PHM. The government should become more active by engaging with users and groups online, answering questions, and responding to posts. It should also do more to support private actors that confront online PHM, by providing them information, institutional guidance, and possible recognition and funding. But government speech alone is insufficient. More active

---

[156] *But see id.* at 1814–21.

[157] *See, e.g.*, Rina Raphael, *TikTok is Flooded with Health Myths. These Creators are Pushing Back.*, N.Y. Times (June 29, 2022), https://www.nytimes.com/2022/06/29/well/live/tiktok-misinformation.html [https://perma.cc/JL57-BFRD] ("For every large creator who is genuinely evidence-based, you've got 50 or 60 big creators who spread misinformation.").

[158] Aimei Yang, Jieun Shin, Alvin Zho, Ke M. Huang-Isherwood, Eugene Lee, Chuqing Dong, Hye Min Kim, Yafei Zhang, Jingyi Sun, Yiqi Li, Yuanfeixue Nan, Lichen Zhen & Wenlin Liu, *The Battleground of COVID-19 Vaccine Misinformation on Facebook: Fact Checkers Vs. Misinformation Spreaders*, 2 Harv. Kennedy School Misinformation Rev. 1, 2 (2021).

[159] United States v. Alvarez, 567 U.S. 709, 727 (2012).

[160] Abrams v. United States, 250 U.S. 616, 630 (1919) (commenting on the assumption that truth will emerge out of the marketplace of ideas: "That at any rate is the theory. . . . It is an experiment, as all life is an experiment"). *See also* Edward Glaeser & Cass R. Sunstein, *Does More Speech Correct Falsehoods?*, 43 J. Legal Stud. 65, 69–71 (2014); R.H. Coase, *The Market for Goods and the Market for Ideas*, 64 Am. Econ. Rev. 384, 390 (1974); Paul H. Brietzke, *How and Why the Marketplace of Ideas Fails*, 31 Val. U. L. Rev. 951, 953–954 (1997).

[161] *See generally* Ari Ezra Waldman, *The Marketplace of Fake News*, 20 U. Pa. J. Const. L. 845 (2018); Dawn Carla Nunziato, *Contemporary Free Speech: The Marketplace of Ideas a Century Later*, 94 Notre Dame L. Rev. 1519 (2019); Alexander Tsesis, *Contemporary Free Speech: The Marketplace of Ideas a Century Later*, 94 Notre Dame L. Rev. 1585 (2019); Emily A. Thorson & Stephan Stohler, *Maladies in the Misinformation Marketplace Essays*, 16 First Amend. L. Rev. 442 (2017).

measures ought to complement such efforts, if governments have any chance to really confront PHM.

C. Consumer Protection Law

Federal and state consumer protection laws protect consumers against unfair trade and credit practices involving faulty and dangerous goods or dishonest claims or tactics.[162] These laws are well-equipped to deal with scammers who disseminate PHM to defraud consumers. Recently, the Federal Trade Commission ("FTC") has noted "a surge in consumer complaints stemming from a broad range of deceptive Covid-related schemes."[163] The FTC responded by sending hundreds of warning letters to sellers who (falsely) claimed that their products can treat or prevent COVID-19, requiring them to stop.[164] Additionally, the Federal Drug Administration ("FDA") has issued hundreds of warning letters to firms for selling fraudulent products that allegedly prevent, treat, mitigate, diagnose or cure COVID-19.[165]

The First Amendment's protection of commercial speech does not prohibit these actions. Government may regulate both factually false commercial advertising and deceptive or misleading commercial advertising, notwithstanding First Amendment protections.[166] But consumer protection law can only go so far. By design, it does not play a role in regulating the false speech of private citizens in non-commercial settings. In other words, consumer protection laws can only regulate PHM disseminated by someone engaged in a commercial transaction. These laws provide no recourse if the same actors disseminate online PHM absent any commercial activity, reaching a very large audience via social media platforms.

Amongst those exploiting this gap, the case of osteopathic physician Joseph Mercola stands out. Mercola was officially warned by the FDA for making false claims about the benefits of his products and other medical

---

[162] The FTC's mission is to protect consumers from "unfair or deceptive acts or practices." Federal Trade Commission Act, 15 U.S.C. § 45(a)(1). State consumer protection laws or "mini-FTC Acts" are modeled on the FTC Act. *See*, *e.g.*, Mass. Gen. Laws ch. 93A, § 2 (2008).

[163] *FTC Outlines Aggressive Approach to Policing Against Pandemic Predators in Testimony Before Senate Commerce Subcommittee*, Fed. Trade Comm'n (Feb. 1, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/02/ftc-outlines-aggressive-approach-policing-against-pandemic-predators-testimony-senate-commerce [https://perma.cc/7NFC-7WFK] (noting more than 292,000 reports associated with COVID-19 frauds in the two-year period ending January 2022, reflecting $674 million in fraud losses).

[164] Fed. Trade Comm'n, Staff Report of the Federal Trade Commission 4–7 (2021) (relying on FTC Act and expanded authority under the COVID-19 Consumer Protection Act of 2020 to justify warning letters).

[165] Office of Regulatory Affairs, *Fraudulent Coronavirus Disease 2019 (COVID-19) Products*, Food and Drug Admin., https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-COVID-19-products [https://perma.cc/7EVJ-FL4W].

[166] *See, e.g.*, Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, 425 U.S. 748, 771 (1976) (allowing commercial speech restrictions that are content-neutral, serve a significant government interest, and leave ample alternative communication channels).

procedures, and had to refund nearly $2.6 million to the FTC for deceptive claims about tanning beds reducing risks of skin cancer.[167] More recently, the FDA issued a warning letter to Mercola regarding his sale of unapproved and misbranded products related to COVID-19.[168] Despite these regulatory actions, Mercola has reportedly made over one hundred million dollars in the past few decades largely from the sale of natural health products (including vitamin supplements, some of which he claims are alternatives to vaccines) and has been actively spreading PHM from which he directly benefits.[169] Mercola is a key figure in what the Center for Countering Digital Hate (CCDH) has dubbed the "Disinformation Dozen"—that is, the twelve anti-vaccination activists who have been most influential in spreading anti-vaccine messaging through social media.[170] He earned this dubious distinction by publishing over 600 anti-vaccination articles on Facebook, with a single article reaching over 400,000 people,[171] and his combined personal social media accounts across major social media platforms reach around 3.6 million followers.[172] However, except when Mercola also engages in a commercial transaction, consumer protection law has no authority to rein in these activities.

D. Medical Malpractice and Board Disciplinary Actions

Both federal and state governments regulate health professionals' speech via licensing requirements, limits on advertising, and medical malpractice liability.[173] In the early days of the COVID-19 pandemic, many front-line doctors and nurses were potentially exposed to such liability, simply because of the sheer prevalence of the disease and the lack of scientific consensus on its cause, treatment, or cure.[174] To solve this problem, the U.S. Secretary of Health and Human Services ("HHS") issued a letter urging all

---

[167] Neena Satija & Lena H. Sun, *A Major Funder of the Anti-vaccine Movement Has Made Millions Selling Natural Health Products*, WASH. POST (Dec. 20, 2019), https://www.washingtonpost.com/investigations/2019/10/15/fdc01078-c29c-11e9-b5e4-54aa56d5b7ce_story.html [https://perma.cc/SKF7-BTZ7].

[168] Letter from William A. Correll, Director, Office of Compliance, Center for Food Safety and Applied Nutrition, to Dr. Joseph M. Mercola, Mercola.com, LLC (Feb. 18, 2021), https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/mercolacom-llc-607133-02182021 [https://perma.cc/RE8Y-2BRJ].

[169] Satija & Sun, *supra* note 167.

[170] Sheera Frenkel, *The Most Influential Spreader of Coronavirus Misinformation Online*, N.Y. TIMES (July 24, 2021), https://www.nytimes.com/2021/07/24/technology/joseph-mercola-coronavirus-misinformation-online.html [https://perma.cc/WHM3-N4UU].

[171] *Id.*

[172] *See* CTR. FOR COUNTERING DIGIT. HATE, THE DISINFORMATION DOZEN: WHY PLATFORMS MUST ACT ON TWELVE LEADING ONLINE ANTI-VAXXERS 7 (2021) (listing Mercola in the number one spot).

[173] *See* Claudia E. Haupt, *Professional Speech*, 125 YALE L.J. 1238, 1240 (2016).

[174] *See* Benjamin J. McMichael, John R. Lowry, William H. Frist & R. Lawrence Van Horn, *COVID-19 and State Medical Liability Immunity*, HEALTH AFFS. (May 14, 2020), https://www.healthaffairs.org/do/10.1377/forefront.20200508.885890/full/ [https://perma.cc/M4GD-DVPB].

state governors to provide civil immunity from medical liability for healthcare professionals treating COVID-19.[175] Accordingly, several state governors and state legislatures ordered and enacted immunity for providers effective immediately upon the declaration of a public health emergency.[176] But even if, following the emergency, medical malpractice lawsuits were to resume and in large numbers, they are unlikely to have much impact on the spread of PHM. To hold healthcare professionals liable for malpractice, plaintiffs must establish a duty of care, failure to meet accepted standards of medical care, causation, and damage.[177] Medical malpractice cases focus on an accepted standard of care regarding a specific individual and hence on a physician's professional advice within the doctor-patient relationship. What physicians say on social media or TV or radio talk shows, however, is of no concern in a medical malpractice action. As Haupt observes, "speech by a professional outside of the professional-client relationship is not professional speech."[178] As such, it is robustly protected under the First Amendment, even if it departs from professional wisdom.[179] As Haupt aptly expresses the point: "a professional may give bad advice to millions of viewers—but not to one client."[180] Thus, as long as health professionals meet the relevant standard of care and act professionally, their professional speech is protected.[181] Consequently, medical malpractice liability is not a useful tool for addressing PHM.

In theory, disciplinary actions by medical boards against licensed professionals who promote PHM seem like a promising tool. In July 2021, the Board of Directors of the Federation of State Medical Boards ("FSMB") warned that "[p]hysicians who generate and spread COVID-19 vaccine misinformation or disinformation are risking disciplinary action by state medical boards, including the suspension or revocation of their medical license."[182] The FSMB noted that physicians have an "ethical and profes-

---

[175] Letter from Alex M. Azar II, Secretary of Health and Human Services, to Governors (March 24, 2020), https://www.nga.org/wp-content/uploads/2020/03/Governor-Letter-from-Azar-March-24.pdf [https://perma.cc/QFK7-MFRP].

[176] McMichael, *supra* note 174; *cf.* The Public Readiness and Emergency Preparedness (PREP) Act, 42 U.S.C. §§ 247d-6d, 247d-6e (providing immunity against losses that arise due to administration or use of the vaccine).

[177] *See* B. Sonny Bal, *An Introduction to Medical Malpractice in the United States*, 467 CLIN. ORTHOPAEDICS & RELATED RSCH. 339, 342 (2009).

[178] Haupt, *Unprofessional Advice*, *supra* note 30, at 681.

[179] *See* Pickup v. Brown, 728 F.3d 1042, 1054 (9th Cir. 2014) ("Thus, outside the doctor-patient relationship, doctors are constitutionally equivalent to soapbox orators and pamphleteers, and their speech receives robust protection under the First Amendment.").

[180] Haupt, *Unprofessional Advice*, *supra* note 30, at 681; *see generally* POST, *supra* note 31.

[181] Haupt, *Professional Speech*, *supra* note 173, at 1267. Conversely, if they fail to meet this standard or act professionally, the First Amendment offers no protection against malpractice liability.

[182] *FSMB: Spreading COVID-19 Vaccine Misinformation May Put Medical License at Risk*, FED'N OF STATE MED. BDS. (July 29, 2021), https://www.fsmb.org/advocacy/news-releases/fsmb-spreading-COVID-19-vaccine-misinformation-may-put-medical-license-at-risk/ [https://perma.cc/Q9TL-L6HY].

sional responsibility" to act for the benefit of patients, and to share information that is "factual, scientifically grounded and consensus-driven for the betterment of public health."[183] Medical boards in several states have adopted FSMB's policy statement, and twelve boards even took action against licensed physicians (as of early 2022).[184] However, medical boards lack resources to monitor physicians' actions on social media unless they are prompted by the filing of a complaint against an individual physician.[185] Additionally, the FSMB statement spawned a political backlash at the state level. This has ranged from reported harassment and intimidation of a health-care worker who alerted the Maryland medical board about the anti-vaccine activity of controversial scientist Robert Malone,[186] to legislative repercussions in dozens of states where bills are under consideration that would limit state medical boards' ability to investigate and act against professionals who spread PHM.[187] In short, the absence of oversight by medical boards allows licensed physicians to trade on their professional credentials while spreading PHM to a large social media audience without much threat of malpractice actions or disciplinary sanctions.

---

[183] *Id.*; *see also Statement About ABEM-Certified Physicians Providing Misleading and Inaccurate Information to the Public*, AM. BD. OF EMERGENCY MED. (Aug. 26, 2021), https://www.abem.org/public/news-events/news/2021/08/27/abem-statement-about-abem-certified-physicians-providing-misleading-and-inaccurate-information-to-the-public [https://perma.cc/YK9R-D6NU] (warning that "making public statements that are directly contrary to prevailing medical evidence can constitute unprofessional conduct and may be subject to review by ABEM").

[184] *See* Blake Farmer, *As State Medical Boards Try to Stamp Out COVID Misinformation, Some in GOP Push Back*, NAT'L PUB. RADIO (Feb. 14, 2022), https://www.npr.org/sections/health-shots/2022/02/14/1077689734/as-state-medical-boards-try-to-stamp-out-covid-misinformation-some-in-gop-push-b [https://perma.cc/K2R2-B3SQ].

[185] *See* Geoff Brumfiel, *This Doctor Spread False Information About COVID. She Still Kept Her Medical License,* NAT'L PUB. RADIO (Sept. 14, 2021), https://www.npr.org/sections/health-shots/2021/09/14/1035915598/doctors-covid-misinformation-medical-license [https://perma.cc/5YEV-F8EJ] (noting that 15 out of 16 licensed physicians promoting misinformation online avoided professional censure and had active licenses in good standing, including Dr. Simone Gold, an emergency physician who spent a year spreading misinformation about the pandemic but had no complaints, disciplinary actions, or malpractice lawsuits on her record); s*ee also* Davey Alba, *The Latest Covid Misinformation Star Says He Invented the Vaccines*, N.Y. TIMES (Apr. 3, 2022), https://www.nytimes.com/2022/04/03/technology/robert-malone-covid.html [https://perma.cc/CGM4-8FYW].

[186] *See* Catherine Offord, *Robert Malone Targets Physician Who Alerted Medical Board to Misinformation*, SCIENTIST (Feb. 19, 2022), https://www.the-scientist.com/news-opinion/robert-malone-targets-physician-who-alerted-medical-board-to-misinformation-69719 [https://perma.cc/5UDT-VEYL].

[187] *See* Press Release, Physicians for Human Rights, COVID-19 Dis-/Misinformation and State Legislature Attacks on Medical Boards Undermine Public Health (Mar. 1, 2022), https://phr.org/news/COVID-19-dis-misinformation-and-state-legislature-attacks-on-medical-boards-undermine-public-health-phr/ [https://perma.cc/642W-2L9W]; Darius Tahir, *Medical Boards Get Pushback As They Try to Punish Doctors for COVID Misinformation*, POLITICO (Feb. 1, 2022), https://www.politico.com/news/2022/02/01/covid-misinfo-docs-vaccines-00003383 [https://perma.cc/Z7YX-HDNR].

E. Negligent Misrepresentation and False Statements

Presumably, spreading online PHM can result in tort liability for negligent misrepresentation that causes bodily harm. Winning a negligent misrepresentation case in this context is challenging. The four elements of the tort are duty of care, negligent misrepresentation, reasonable reliance on the representation, and such reliance physically harming the plaintiff or a foreseeable third party.[188] As with medical malpractice claims, the lack of a duty of care is often fatal to plaintiffs bringing actions for negligent misrepresentation in cases involving the safety or dangers of procedures or treatments in general.[189] It is even more difficult to establish a duty of care for anti-vaccine activists who are not even health professionals, and thus lack any special relationship with their audience that would create such a duty of care. Nor does the blogosphere have any reasonable basis to rely on medical advice from activists, celebrities, politicians, and other non-professionals lacking medical expertise as opposed to following the advice of their own health providers.[190] Scholars have argued that liability for misrepresentation can be applied in unique circumstances of spreading PHM.[191] But in most cases, the First Amendment imposes limits on the misrepresentation tort, especially in the context of information published to the general public.[192] Hence, general publications of PHM on platforms are likely immune from such misrepresentation claims.

F. Regulating False Speech

As hinted at previously, speech regulation usually raises First Amendment concerns. But what about misinformation? Are false statements protected by the First Amendment even when they cause serious harms to those who rely on them? Seemingly, the answer is yes. In *United States v. Alvarez*, the Supreme Court held that content-based restrictions on false statements are invalid.[193] In his opinion for the plurality, Justice Kennedy identified a few "traditional categories" of permissible content-based regulation (includ-

---

[188] *See* Reiss & Diamond, *supra* note 17, at 534–35.

[189] *See, e.g.,* Bailey v. Huggins Diagnostic & Rehab. Ctr., Inc., 952 P.2d 768, 772 (Colo. App. 1997) (finding no duty between a dentist who made public claims about the dangers of amalgams and a patient relying on such representations to replace her amalgams with an inferior substance).

[190] *See* Reiss & Diamond, *supra* note 17, at 534–44 (discussing activists targeting this community).

[191] *Id.*

[192] *Id.* at 574 ("Duty is weakest, and the First Amendment is strongest, where the [anti-vaccine] information is posted in a public forum for general consumption.").

[193] *See generally* United States v. Alvarez, 567 U.S. 709 (2012) (overturning a federal statute, the Stolen Valor Act (SVA), which criminalized false claims about the receipt of military decorations or medals, and holding in broad terms that content-based restrictions on false statements are invalid). For more on *Alvarez* see generally HASEN, *supra* note 128; CASS R. SUNSTEIN, LIARS: FALSEHOODS AND FREE SPEECH IN AN AGE OF DECEPTION (2021).

ing obscenity, speech integral to criminal conduct, and fraud) but declined to create a new categorical exclusion for false statements as such.[194] Importantly he declined to create a new categorical exclusion from First Amendment protection for false statements, especially when there is no evidence of fraud or some other legally cognizable harm associated with falsity.[195] The Court applied strict scrutiny and found that the relevant regulation failed to meet the burden.[196] Notably for our purposes, the Court was not persuaded that a direct causal link existed between the restriction imposed and the injury to be prevented.[197] Nor was it persuaded that counter-speech would not suffice to achieve the state interest, echoing the marketplace of ideas rationale.[198]

However, there are two possible distinctions between the law at issue in *Alvarez* and possible laws regulating PHM. First, it's unclear that *Alvarez*'s reliance on counter-speech and the marketplace of ideas is apt for PHM. The false speech in *Alvarez* was limited in scope and easily verifiable to all listeners: Alvarez falsely introduced himself at a public meeting of the local water board as a retired Marine, who had been wounded and awarded the Medal of Honor.[199] His lies were quickly exposed, subjecting him to public ridicule online and in the local press.[200] All those make *Alvarez* an easy case for showing the efficacy of counter-speech and the marketplace of ideas.[201] However, when it comes to online PHM, the spread of the message is unimaginably broader and the practical ability to counter it is challenging. Hence, the effectiveness of counter-speech with regard to PHM is doubtful at best.[202] Second, both the plurality and concurring opinions in *Alvarez* con-

---

[194] *Alvarez*, 657 U.S. at 717–18 (plurality opinion).

[195] *Id.* at 722–23.

[196] *Id.* at 724.

[197] *Id.* at 726 ("The Government points to no evidence to support its claim that the public's general perception of military awards is diluted by false claims such as those made by Alvarez.").

[198] *Id.* at 727. Note that both the concurring and dissenting opinions also invoked the marketplace of ideas rationale. *See id.* at 732 (Breyer, J., concurring); *id*. at 746 (Alito, J., dissenting).

[199] *Id.* at 713–14.

[200] *Id.* at 726–27.

[201] Indeed, the Court relied on this justification explicitly. *See id*. at 727 (Kennedy J., plurality); *id.* at 732 (Breyer, J., concurring); *id.* at 746 (Alito, J., dissenting). For criticism of this view, see, e.g., James Weinstein, *What Lies Ahead?: The Marketplace of Ideas,* Alvarez v. United States*, and First Amendment Protection of Knowing Falsehoods*, 51 SETON HALL L. REV. 125, 136 (2020) (noting that this rationale has been "trenchantly criticized in the scholarly literature").

[202] For recent studies of how anti-vaccination views spread online and why they prevail over pro-vaccination views, see generally Neil F. Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty & Yonatan Lupu, *The Online Competition Between Pro- and Anti-Vaccination Views*, 582 NATURE 230 (2020) (describing the emergence of anti-vaccine clusters among 100 million Facebook users and the features of this cluster that explain why negative views have become "so robust and resilient"); Barbara P. Billauer, *Muzzling Anti-Vaxxer FEAR Speech: Overcoming Free Speech Obstacles with Compelled Speech*, 76 U. MIAMI L. REV. 1, 55 (2021) (explaining that in social media context, counter-speech risks "what is

ceded that false speech may be regulated in laws that address some recognized harm, including defamation, obscenity, perjury, impersonation of public officials, and so on.[203] As discussed above, PHM poses a recognized social harm that government regularly addresses as a compelling state interest. Hence, laws and regulations that confront PHM support public health and address a cognizable harm, and thus are distinguishable from *Alvarez*.[204]

In any case, as long as the *Alvarez* decision stands, any broadly worded law restricting ordinary, private citizens (i.e., non-commercial or non-professional actors) from producing, receiving, or sharing PHM would likely not survive First Amendment review.

G. Social Media Platforms and Section 230

All the regulatory approaches considered above seek to hold the speaker liable for online PHM, with limited success. What about holding platforms accountable instead? After all, social media platforms are among the most important channels for spreading PHM. And Facebook, Twitter, YouTube, and other major platforms disseminate misinformation at high velocity, use algorithms to target interested recipients, and amplify content most likely to generate engagement. Thereby, they extend the reach of PHM to huge and receptive audiences. So why not hold them accountable for online PHM?

The short answer is that Section 230 of the Communications Decency Act of 1996 blocks this move by immunizing "interactive computer services" against civil liability for publishing third-party content or for the removal of content under certain circumstances.[205] Congress enacted the statute to promote private ordering and to enable early online services to take down offensive content without exposing themselves to publisher's liability.[206] Courts adopted a broad interpretation of Section 230, interpreting "interactive computer services" to cover new social media platforms like Facebook or Twitter.[207] Courts also held that the statute immunizes platforms from liability as publishers of another's information, subject to a few statu-

known as the 'back-fire' effect, where the concern that repeating false information, even to correct it, can strengthen beliefs in" unscientific myths). *See also* Richard L. Hasen, *Cheap Speech and What It Has Done (to American Democracy)*, 16 FIRST AMEND. L. REV. 200 (2018) (arguing that "cheap speech" exacerbates polarization and that counter-speech may not be enough to deal with it); Toni Marie Massaro & Helen L. Norton, *Free Speech and Democracy: A Primer for 21st Century Reformers* 54 U.C. DAVIS L. REV. 1631, 1645 (2021) (noting that in the online setting, counter-speech is not a "realistic option for those without the resources or expertise to confront well-aimed lies with rebuttals of equal volume, speed, and listener-targeted precision").

[203] *Alvarez*, 567 U.S. at 717–18 (plurality opinion); *id.* at 731–32, 734–37 (Breyer, J., concurring).

[204] *See supra* Section I.B.

[205] 45 U.S.C. § 230.

[206] *See* 45 U.S.C. § 230(b).

[207] *See, e.g.*, Klayman v. Zuckerberg, 753 F.3d 1354, 1358 (D.C. Cir. 2014) (Facebook); Fields v. Twitter, Inc., 217 F. Supp. 3d 1116, 1121 (N.D. Cal. 2016) (Twitter).

tory exceptions,[208] and readily immunized platforms for removing content in good faith based on their community guidelines. For instance, Facebook defeated a suit by Children's Health Defense ("CHD"), the anti-vaccination organization founded and run by Robert Kennedy Jr., alleging that Facebook's content moderation decisions amounted to "censorship."[209] In rare cases, courts restricted Section 230 immunity based on a finding that a platform "is responsible in whole or in part, for the creation or development of the information."[210]

To illustrate the breadth of the immunity that Section 230 provides to platforms, consider *Dyroff v. The Ultimate Software Group, Inc.*[211] Dyroff sued Ultimate Software for its alleged role in the death of her son, Wesley Greer, who posted a message asking where he could buy heroin in Jacksonville. Greer then received a notification from Ultimate's website, indicating that another user had responded to his question, and Greer contacted this user, purchased fentanyl-laced heroin from him, and died the next day.[212] The Ninth Circuit held that Section 230 immunized Ultimate Software from liability and barred Dyroff's claims.[213] Courts recently upheld this broad interpretation of Section 230 in other cases involving recommendation algorithms. In *Force v. Facebook*[214] and *Gonzalez v. Google*,[215] the Second Circuit and the Ninth Circuit, respectively, relied on *Dyroff* and other cases to extend Section 230 immunity to platforms' recommendations of content that related to terrorist organizations and activities. Hence, platforms will likely be successful in invoking Section 230 immunity against alleged liability for using algorithmic recommendations in spreading PHM, even if it can be shown that such PHM led to individual harms or degraded public health.

Having said that, the Supreme Court recently granted certiorari in *Gonzalez* and is set to discuss the breadth of Section 230 immunities, specifically with regard to personalized algorithmic recommendations (or "targeted rec-

---

[208] Section 230(e) expressly provides that the Section 230 safe harbor will not apply to: (1) federal criminal laws; (2) intellectual property laws; (3) any state law that is "consistent with" Section 230; (4) the Electronic Communications Privacy Act of 1986; and (5) certain civil actions or state prosecutions where the underlying conduct violates specified federal laws prohibiting sex trafficking. 45 U.S.C. § 230(e).

[209] *See* Children's Health Def. v. Facebook, Inc., 546 F. Supp. 3d 909, 915, 945 (N.D. Cal. 2021). The content moderation decisions included marking CHD content as false based on independent fact-checking, disabling CHD's ability to dispute Facebook's content moderation decisions, deactivating the "donate" button on CHD's pages, blocking CHD from placing ads, and eventually placing a warning label on CHD's page. *Id.* at 919–21.

[210] *See, e.g.,* Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1167–68 (9th Cir. 2008) (precluding Section 230 immunity where defendant acted as a co-developer of content by "materially contributing to its alleged unlawfulness").

[211] 934 F.3d 1093 (9th Cir. 2019).

[212] *Id.* at 1095–96.

[213] *Id.* at 1097, *cert. denied*, 140 S. Ct. 2761 (2020).

[214] 934 F.3d 53, 64–71 (2d Cir. 2019).

[215] 2 F.4th 871, 890–97 (9th Cir. 2021).

ommendations" in the Court's language).[216] The Court seems poised to use this case to poke holes in the broad protections that Section 230 provides to platforms.[217] We think that a generic approach that analyzes all uses of recommendation algorithms as one is mistaken. A more nuanced approach that examines the uses of recommendation algorithms in specific platforms and contexts is advisable. Moreover, we think there are better ways to address the many problems that Section 230 invokes for online speech.[218] Time will tell what *Gonzalez* will bring. Anyhow, at the time of writing, Section 230 immunity blocks lawsuits related to the posting, distribution, and amplification of PHM.

Some legislative efforts try to avoid this outcome by carving out exceptions to Section 230 for disseminating PHM. These include dozens of bills to fund public awareness campaigns to dispel misinformation about COVID-19 symptoms, testing, or treatment;[219] dozens of bills to amend Section 230;[220] and dozens of bills that would directly regulate social media platforms' use of algorithms.[221] Modifying Section 230 is ineffective, however, when liability is lacking even without Section 230 immunity. Consider Senator Amy Klobuchar's S. 2448, the Health Misinformation Act.[222] It suggests that a social media platform would lose its immunity if it algorithmically promotes health misinformation during public health emergencies, as defined by HHS. However, loss of immunity is an empty threat when, as we have seen, there are no laws prohibiting PHM.[223] In other words, in the absence of any cause of action exposing a publisher of health misinformation to potential liability, amending Section 230 to withdraw immunity accomplishes nothing.

---

[216] Question Presented, Gonzalez v. Google LLC, No. 21-1333, (U.S. Oct. 3, 2022) https://www.supremecourt.gov/docket/docketfiles/html/qp/21-01333qp.pdf [https://perma.cc/D7JA-WJ7P].

[217] *See* Tomer Kenneth & Ira Rubinstein, *Gonzalez v. Google: The Case for Protecting "Targeted Recommendations,"* 72 DUKE L.J. ONLINE (forthcoming Apr. 2023).

[218] *Id*.

[219] Gallo & Cho, *supra* note 36 at tbl. B-2.

[220] *Id.* at tbl. B-1.

[221] *See* Spandana Singh, *Regulating Platform Algorithms: Approaches for E.U. and U.S. Policymakers*, NEW AMERICA (Dec. 1, 2021), https://www.newamerica.org/oti/briefs/regulating-platform-algorithms/ [https://perma.cc/JN4V-5ETK]. State lawmakers have also been active in calling for legislation on the use of algorithms. *See generally* NAT'L CONF. OF STATE LEGISLATURES, *Legislation Related to Artificial Intelligence* (Aug. 26, 2022), https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx [https://perma.cc/L62P-CT6C].

[222] S. 2448, 117th Cong. (2021).

[223] *See* Mark MacCarthy, *Senator Amy Klobuchar Seeks to Quell Health Misinformation on Social Media*, BROOKINGS (July 27, 2021), https://www.brookings.edu/blog/techtank/2021/07/27/senator-amy-klobuchar-seeks-to-quell-health-misinformation-on-social-media/ [https://perma.cc/ZG8X-YBB7].

V. New Paths for Government Action Against Public Health Misinformation

So far, we have established that online PHM is a considerable problem, that the solution cannot and should not be left only to platforms, and that existing laws are insufficient. We now turn to our positive argument. In this Part, we discuss several paths for governments to confront online PHM. Exploring these varied approaches will illustrate, against the common understanding, that the government has considerable power to counteract PHM. Specifically, we explain how the government can use previously untapped sources, such as soft regulation and fresh thinking about legislative reforms, to reach those goals. These solutions are not perfect. But they have two considerable advantages. They directly confront online PHM by influencing the main arena for disseminating it—social media platforms. And these solutions are also feasible under existing legal doctrines. Thus, the solutions examined below fare much better than the existing legal solutions discussed in Part IV.

A. Soft Regulation

The government has many ways to influence the circulation of online PHM. In this section, we survey several "softer" avenues of influence that government actors may use to confront PHM. This section focuses on methods that allow democratic governments to influence platforms without directly controlling platforms' policies and their implementation or severely threatening platforms' independence as private entities.[224]

Soft regulation of PHM will often be directed at what Jack Balkin calls the "infrastructure of freedom of expression," a "technological and regulatory infrastructure . . . [that is] produced through government regulation, through government subsidies and entitlement programs, and through technological design."[225] In the digital age, such infrastructure is mostly privately held, and it includes platforms that distribute and feature speech like social networks or search engines, as well as domain-name systems, internet protocols, and network and broadband providers.[226] When the infrastructure is privately owned (and as centralized as in online speech), government attempts to regulate speech focus on that infrastructure.[227] The government

---

[224] More robust forms of influence that still fall short of direct regulation are beyond the scope of this Article. For more information, see generally Gary King, Jennifer Pan & Margaret E. Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument*, 111 Am. Pol. Sci. Rev. 484 (2017) (discussing Chinese government control over platforms' decisions through state involvement in private ownership).

[225] Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society Commentary*, 79 N.Y.U. L. Rev. 1, 48 (2004).

[226] Balkin, *New School*, *supra* note 150, at 2303–04.

[227] Governments also have other ways to use platforms to promote their interests. *See, e.g.*, Tomer Kenneth, *Personalization of Smart-Devices: Between Users, Operators, and*

often targets this infrastructure in order to govern or regulate speech, because such infrastructure is a bottleneck that facilitates control over millions of users that are otherwise harder to reach.[228]

Soft regulation involves enlisting private actors that control this infrastructure to promote the government's interests.[229] Balkin identified three kinds of speech regulation relevant for digital infrastructures: collateral censorship, prior restraint, and public-private cooperation.[230] We focus on the latter. Public-private cooperation refers to measures that governments use to influence platforms.[231] They encourage platforms to adopt or apply specific regulations or drive platforms to share access to data they collect.[232]

Platforms have a lot to gain from this cooperation. Consider two examples. First, platforms have interests in moderating, cultivating, and promoting or suppressing particular content.[233] These interests can arise from a sense of corporate responsibility or from economic reasons, aiming to capture more users' attention.[234] Simply stated, ISIS beheadings, pornography, and PHM, might be bad business for platforms like Facebook and Twitter that cater to the mainstream.[235] Still, moderating content is hard. Devising specific content moderation schemes—deciding which kinds of content is problematic and striking the right balance between those interests and free speech—requires considerable effort. And such decisions often expose platforms to public criticism and scrutiny.[236] Also, even if platforms have perfect normative intentions, such efforts might cut against their business model, which affects outcomes.[237] Cooperating with the government on content moderation decisions shifts some of these problems to the state. Platforms can justify removing some content based on compliance with government's

---

Prime-Operators, 70 DEPAUL L. REV. 497, 517–20 (2021) (discussing how regulators can use smart-devices to enforce and create more personalized regulation).

[228] Balkin, *New School*, *supra* note 150, at 2303–04; *see also* Molly K. Land, *Against Privatized Censorship: Proposals for Responsible Delegation*, 60 VA. J. INT'L L. 363, 374 (2019) ("The sheer volume of content available online, combined with the challenge of identifying the source of such content and the difficulty of pursuing the actual violators across borders, makes policing online speech through traditional means costly and cumbersome.").

[229] Balkin, *New School*, *supra* note 150, at 2305.

[230] *Id.* at 2308–29; *see also* Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2015–21 (2018).

[231] *See* Balkin, *New-School*, *supra* note 150, at 2324–29; Michael D. Birnhack & Niva Elkin-Koren, *The Invisible Handshake: The Reemergence of the State in the Digital Environment*, 8 VA. J.L. & TECH. 6, 35–44, 122–42 (2003).

[232] Balkin, *New-School*, *supra* note 150, at 2324–29; *see also* Birnhack & Elkin-Koren, *supra* note 231, at 122–42.

[233] Liu et al., *supra* note 95, at 25 ("[P]latforms are more eager than a social planner to conduct content moderation motivated by their own self-interest.").

[234] Klonick, *supra* note 107, at 1625–30; Balkin, *Free Speech Is a Triangle*, *supra* note 230, at 2022–23.

[235] That is why platforms like Facebook adopt complicated "community standards" to secure the "integrity" of the platform. *See Facebook Community Standards*, META TRANSPARENCY CENTER, https://transparency.fb.com/policies/community-standards/ [https://perma.cc/S4K9-VH43].

[236] Klonick, *supra* note 107, at 1631–35.

[237] *See supra* Part III.

guidelines or upon legal order by the state. This approach might be a lot easier than trying to justify a general speech policy or publicly defend a specific decision based on its merits. So, by deferring to governments on speech regulation, platforms can shift the burden of drawing controversial lines and shift the blame of possible outrage from those decisions.

Second, when platforms and the state agree that some content is undesirable, the state's coercive powers can help disincentivize its publication on platforms ex-ante. When the state bans the distribution of some content and holds the distributor (speaker) liable, it disincentivizes dissemination of this content.[238] A platforms' cooperation with the state (and perhaps merely the inclination to cooperate) on those issues also may disincentivize publication. For instance, platforms can publicly agree to share with the state metadata and other possibly identifying information about users that uploaded child pornography or hate speech.[239] In doing so, platforms signal that they will help the state find those users. In turn, this increases the speaker's chances of being caught (and therefore punished) by the state, making it riskier for them to publish this content on that platform. Thus, cooperation with the state allows platforms to use governments' coercive power to disincentivize unwanted content.

Soft regulation seems especially appropriate for confronting online PHM. Communicating public health information and confronting misinformation are crucial public health measures,[240] and the most relevant speech arena is controlled by private platforms.[241] This approach is also better than traditional governmental efforts like passing laws and regulations. Soft regulation recognizes the inadequacy of traditional regulation in responding to online speech[242] and the fundamental changes that have occurred in the infrastructure of freedom of speech.[243] Moreover, it works. The European Commission concluded that self-regulation is an effective measure to regulate online speech.[244] It drives platforms to increase their monitoring and review-

---

[238] For instance, imagine that WhatsApp identified and notified the police when its users shared non-consensual (or child) pornographic photos and videos. That would disincentivize users from sharing this content (at least on this platform). *See* Kashmir Hill, *A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal*, N.Y. TIMES (Aug. 21, 2022), https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html [https://perma.cc/XT24-85XN] (describing how Google notified authorities about existence of child nudity on a father's phone). *See also* Kenneth, *Personalization of Smart-Devices*, *supra* note 227, at 517–20 (discussing how smart-devices can extend law's practical reach).

[239] *See, e.g.*, Balkin, *New School*, *supra* note 150, at 2325 (discussing collaborative or mandatory ways for platforms to share information with the state).

[240] *See supra* Section I.B.

[241] *See supra* notes 226–30.

[242] *See infra* notes 279, 290.

[243] *See supra* notes 225–29.

[244] *The EU Code of Conduct on Countering Illegal Hate Speech Online*, EUR. COMM'N, https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en [https://perma.cc/GCF2-Y9CH].

ing efforts and to be more diligent in regulating speech.[245] This method also avoids the rigidity and stiffness of most state regulation. Platforms guided by soft regulations would be more inclined to remove limitations and loosen strict measures when conditions allow, compared to strict regulation that often becomes sticky years later.[246] In light of the importance of confronting online PHM for public health,[247] government should harness the immense power of platforms over online speech.

Of course, soft regulation is not flawless. The line between cooperation and state coercion might be muddier than it seems at first blush.[248] An apparently voluntary measure might be complemented with a more restrictive measure (or a credible threat of such) that would make compliance as unavoidable as in standard regulation.[249] For those reasons, scholars criticize these methods as being unidirectional rather than cooperative—the state requires platforms to act in certain ways in exchange for the state's agreement not to use harsher measures against them.[250] According to this criticism, governments (mis)use private ordering to achieve governments' policy preferences,[251] often going beyond what legal measures would allow for direct regulation,[252] thereby using platforms to "launder" policy preferences and unduly censure lawful expression.[253] Those criticisms also raise concerns about the state's overreach and extensive use of private platforms to regulate speech, all under the guise of public-private cooperation.[254]

To the best of our knowledge, these kinds of soft-regulation solutions have not yet been utilized in the United States. While the government has cooperated with and attempted to influence private companies in the past, it has yet to adopt the comprehensive approach that the EU and other countries

---

[245] Council of the European Union, *Assessment of the Code of Conduct on Hate Speech on Line State of Play*, EUR. COMM'N 3–4 (Sep. 27, 2019), https://commission.europa.eu/system/files/2020-03/assessment_of_the_code_of_conduct_on_hate_speech_on_line_-_state_of_play__0.pdf [https://perma.cc/D6NM-ASUH].

[246] On sticky regulations that are difficult to change, see generally Aaron L. Nielson, *Sticky Regulations*, 85 U. CHI. L. REV. 85 (2018). On the relative ease of changing platform regulation, see, e.g., *supra* note 81 and accompanying text (describing Facebook rolling back COVID-19 regulations); *supra* note 101 and accompanying text (describing Twitter revolutionizing its content moderation).

[247] *See supra* Section I.B.

[248] Balkin, *Free Speech Is a Triangle*, *supra* note 234, at 2028–32.

[249] Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 43–57 (2019). *See also* European Commission Press Release IP/21/5082, *EU Code of Conduct against illegal hate speech online: results remain positive but progress slows down* (Oct. 7, 2021) [https://perma.cc/T72L-3SD2] (discussing the need to complement gaps of the hate speech code using the Digital Service Act).

[250] *See, e.g.*, Land, *Against Privatized Censorship*, *supra* note 228, at 380–86 (arguing that intermediaries' safe-harbors are not a form of cooperation because the platforms do not have viable alternatives to complying).

[251] Bloch-Wehba, *Global Platform Governance*, *supra* note 249, at 29–30.

[252] Land, *Against Privatized Censorship*, *supra* note 228, at 378.

[253] Daphne Keller, *Who Do You Sue? State and Platform Hybrid Power Over Online Speech*, AEGIS SERIES PAPER NO. 1902, 3 (2019).

[254] Balkin, *New School*, *supra* note 153.

have. This might be more than a coincidence. It is possible that this approach is simply not in line with the legal culture—this is not how we do things. The fierce public rebuke of the Disinformation Governance Board, a group set up to coordinate *existing* measures to counter disinformation,[255] is a recent illustration of this mentality.[256] Admittedly, the soft-regulation approach might appear foreign to the U.S. legal tradition. To those who share this intuition, the next few pages are an invitation for reflection. Since the problem of online PHM is serious and existing solutions are inapt, it might be time to reconsider traditional views and explore new horizons.

1. Codes of Conduct: "Voluntary" Self-Regulation

Governments can influence platforms to confront online PHM by cultivating self-regulation. On this approach, governments (or inter-governmental organizations) publish 'codes of conduct' that offer guidance on a range of content-management and platform-governance issues.[257] The codes help cultivate cooperation across platforms and between platforms and states, namely by expressly settling contested issues and creating official paths for engagement and cooperation. The scope, specificity, legal standing, and sanctions for non-compliance may vary among different codes. Most codes have some form of ongoing monitoring, often including reports by the platforms and the states assessing platform compliance. These self-assessment reports, alongside reports by the states and governmental organizations, serve multiple functions. They add information and clarity about platforms' actions, thereby helping governments understand platforms' actions and help platforms coordinate amongst themselves. These reports are also used as a bellwether for platforms, indicating whether governments are content with the status quo and what policy changes (if any) the government might be interested in.

Codes of conduct are not as mandatory as laws or regulations. Instead, they rely on voluntary adoption by platforms. Platforms adopt these codes for a number of reasons. First, they infer that governments will enforce

---

[255] *Fact Sheet: DHS Internal Working Group Protects Free Speech and Other Fundamental Rights When Addressing Disinformation That Threatens the Security of the United States*, Dep't of Homeland Sec. (May 2, 2022), https://www.dhs.gov/news/2022/05/02/fact-sheet-dhs-internal-working-group-protects-free-speech-other-fundamental-rights [https://perma.cc/3YR4-PEG5].

[256] *See* Letter from Charles E. Grassley & Josh Hawley, U.S. Senators, to Alejandro N. Mayorkas, Secretary of Homeland Security, 3–5 (June 7, 2022), https://www.grassley.senate.gov/imo/media/doc/grassley_hawley_to_deptofhomelandsecuritydisinformationgovernanceboard.pdf [https://perma.cc/87HE-PNB8] (warning against the Disinformation Governance Board's intention to cooperate with private platforms to confront disinformation).

[257] *See generally* Bloch-Wehba, *Global Platform Governance*, *supra* note 246; Land, *Against Privatized Censorship*, *supra* note 228; Rutschman, *Self-Regulation*, *supra* note 75, at 59–65.

stricter regulations if the codes are not widely adopted.[258] Second, signing up gives platforms a seat at the table, and hence a more direct opportunity to influence the codes compared to legislation.[259] Third, and relatedly, platforms' interests in regulating such speech often align with those of the regulator, and it is therefore beneficial to rely on the regulator.[260] Fourth, there are reputational benefits to signing on to the codes and potential reputational damages for declining to do so.[261] Finally, the codes allow the signatories to share with their business rivals minimal content regulation standards, thereby solving a potential collective action problem.[262]

The European Union championed this method in its online hate speech and disinformation codes of conduct.[263] Beginning in 2016, EU organs published communications and guidelines for platforms about confronting hate-speech.[264] These codes call on platforms to develop "clear and effective processes;" review and remove illegal hate speech on their services within 24 hours of detection; educate users on these issues; make provisions for notice and flagging of violent or hateful content, including developing "trusted reporter"[265] roles; intensify cooperation between platforms on best practices; disclose information to states about those procedures; and improve the communication between states and platforms regarding content classified as hate speech, so that states can "recognise and notify the companies of illegal hate speech online."[266] Each year, the European Commission pub-

---

[258] Interview with Alexandre de Streel, Academic Director, Ctr. on Regul. in Eur., Chair, EU Observatory on Online Platform Economy; *supra* note 249.

[259] *See The Sounding Board's Unanimous Final Opinion on the So-Called Code of Practice*, EUR. COMM'N (Sept. 24, 2018), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54456 [https://perma.cc/K5HH-RV9M] (noting that the code of practice originated from the "Multistakeholder Forum on Disinformation Online," which included the platforms).

[260] *See supra* text accompanying notes 233–39.

[261] *Joint Call for interest to join the Code of Practice on Disinformation*, EUR. COMM'N (July 9, 2021), https://digital-strategy.ec.europa.eu/en/joint-call-interest-join-code-practice-disinformation [https://perma.cc/64U9-ZKY5].

[262] *Id*. The collective action problem we have in mind is one platform's desire to take some measure, but not wanting to act alone because of concerns about public scrutiny or losing users to rivals. Acting in concert may be beneficial in these cases. For a recent example, see Kate Conger, Mike Isaac & Sheera Frenkel, *Twitter and Facebook Lock Trump's Accounts After Violence on Capitol Hill*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2021/01/06/technology/capitol-twitter-facebook-trump.html [https://perma.cc/7EFE-BUQQ].

[263] Sometimes these codes of conduct do not involve government actors at all. *See, e.g.*, *Australian Code of Practice on Disinformation and Misinformation*, DIGITAL INDUSTRY GROUP INC. (Feb. 22, 2021), https://digi.org.au/wp-content/uploads/2021/10/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-WORD-UPDATED-OCTOBER-11-2021.pdf [https://perma.cc/BF54-NYAN] (The code was created by the Digital Industry Group Inc., founded by Apple, eBay, Google, Meta, and Twitter, among others.).

[264] Bloch-Wehba, *supra* note 249, at 43–51.

[265] *See, e.g.*, *About the YouTube Trusted Flagger Program*, GOOGLE, https://support.google.com/youtube/answer/7554338?hl=EN [https://perma.cc/P8RD-APZC].

[266] *Code of Conduct on Countering Illegal Hate Speech Online*, EUR. COMM'N (June 30, 2016), https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985 [https://perma.cc/6QE8-XUWB].

lishes an evaluation of the code of conduct, thereby monitoring the platforms actions and pressuring them to comply with the code.[267]

Closer to our main topic, the EU also published a code of practice on disinformation. The code calls on platforms (and advertisers) to voluntarily adopt self-regulation to confront disinformation.[268] It requires platforms to "invest in technological means to prioritize relevant, authentic, and accurate and authoritative information," improve transparency, "dilute the visibility of disinformation," and "write an annual account of their work to counter disinformation."[269] The code also includes a "best practices annex" that sets out principles for platforms to follow, such as stopping monetization of disinformation, acting against inauthentic users, and creating reporting systems.[270] The monitoring aspect of this code includes requiring platforms to publish a yearly self-assessment report alongside reports by EU actors.[271] In June 2022, the code was substantively revised. The existing code holds over 160 specific commitments and measures.[272] The code builds on the 2018 code of conduct, which was developed in cooperation with the signatories and is complemented by other EU regulation.[273] It calls on platforms to regulate political advertising, de-monetize disinformation, flag harmful and misleading information, create appeal mechanisms, empower users to confront online disinformation, and cooperate with researchers and fact-checkers.[274] It also requires signatories to implement their commitments under the code within 6 months, and devise an elaborate and detailed reporting and monitoring system.[275] The code of conduct was adopted by 34 signatories including major platforms and AdTech giants.[276] As such, it illustrates the ability of codes of conduct, and soft regulation more generally, to facilitate desired change in online speech governance.

---

[267] *See, e.g.*, Didier Reynders, *Countering Illegal Hate Speech Online: 6th Evaluation of the Code of Conduct*, EUR. COMM'N (Oct. 7, 2021), https://commission.europa.eu/system/files/2021-10/factsheet-6th-monitoring-round-of-the-code-of-conduct_october2021_en_1.pdf [https://perma.cc/A8PZ-VBPS] [hereinafter *6th evaluation*].

[268] *See Shaping Europe's Digital Future—The 2022 Code of Practice on Disinformation*, EUR. COMM'N, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation [https://perma.cc/9NQ4-9KPV] (last visited Mar. 3, 2023) [hereinafter *2022 Code of Practice on Disinformation*].

[269] *EU Code of Practice on Disinformation*, EUR. COMM'N, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454 [https://perma.cc/7FZM-SK5C].

[270] *Annex II Current Best Practices from Signatories of the Code of Practice*, EUR. COMM'N , https://eaca.eu/wp-content/uploads/2022/02/annex_best_practices_final_docx_D33C8E56-EDEE-3021-BAFB4B185F82C7ED_54455-1.pdf [https://perma.cc/W7ZS-X49T].

[271] *2022 Code of Practice on Disinformation*, *supra* note 268.

[272] *Id.*

[273] *Id.*; *see, e.g.*, Regulation 2022/2065 of Oct. 19, 2022, Regulation on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) §§ 104, 106 (stating codes of conduct could be a basis for self-regulatory efforts but participating in the code does not presume compliance).

[274] *2022 Code of Practice on Disinformation*, *supra* note 268.

[275] *Id.*

[276] *Id.*

Given the relative success of the guidelines and code of conduct against online hate-speech and disinformation, employing this approach to confront online PHM seems advised. Developing a code of conduct for confronting online PHM has several advantages. It is relatively easy and expedient to adopt and modify such code, especially compared to enacting and amending legislation. Given the fast-changing nature of both public health crises and online speech, speed matters. The benefits of this approach were applied with regard to PHM in the early days of the COVID-19 pandemic. In March 2020 the existing disinformation code was supplemented by a "need for additional efforts," requiring platforms to provide a monthly report "on their actions to promote authoritative content, improve users' awareness, and limit coronavirus disinformation and advertising related to it."[277] Major platforms are complying with this new requirement, and arguably at least some of the synchronization and harmony between the platform's actions to confront online PHM are the result of the guidelines.[278] Additionally, given their voluntary and cooperative nature and the lack of harsh penalties, such codes are more likely to pass muster—legally and socially—compared to full-blown laws.[279]

At the time of writing this paper, we know of no comparable efforts by the U.S. government to confront misinformation. There are less substantive attempts that lack most of the influence levers mentioned above. Those might qualify as soft regulation, but they are far removed from trying to adopt this method in the U.S. on PHM which included a section on what platforms should do to confront it. Those measures included better monitoring of misinformation, detecting PHM from "super-spreaders," prioritizing the protection of health professionals and amplifying communication from trusted sources.[280] More recently, the Surgeon General officially asked the public, and specifically platforms, to share information about PHM and its effect on patients and public health.[281] However, these efforts fall short of the EU guidelines: they are not as robust, there is little indication for cooperation with platforms, and no ongoing mechanisms to learn platforms' abilities and concerns or to signal them what they should do. This is unfortunate. Government should seriously consider using codes of conduct as a way to influence platforms.

---

[277] *Coronavirus: EU Strengthens Action to Tackle Disinformation,* EUR. COMM'N (June 10, 2020), https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1006 [https://perma.cc/U82H-VWSQ].

[278] *See Shaping Europe's Digital Future—Reports on January and February Actions*, EUR. COMM'N (Mar. 31, 2022), https://digital-strategy.ec.europa.eu/en/library/reports-january-and-february-actions [https://perma.cc/8P4C-646Y] (including monthly reports from TikTok, Meta, Twitter, Microsoft, and Google).

[279] Rutschman, *Self-Regulation*, *supra* note 75, at 66–69 (discussing those considerations in support of adopting a code of conduct to confront COVID-19 misinformation).

[280] SG REPORT, *supra* note 32, at 12.

[281] Davey Alba, *The Surgeon General Calls on Big Tech to Turn Over COVID-19 Misinformation Data*, N.Y. TIMES (Mar. 3, 2022), https://www.nytimes.com/2022/03/03/technology/surgeon-general-covid-misinformation.html [https://perma.cc/C4QC-ABBH].

Notably, first signs of change are already on the horizon. For example, the U.S. Department of State published a Declaration for the Future of the Internet, which over sixty countries joined as partners.[282] The Declaration identified the promises of the Internet as a "network of networks" for humanity.[283] It noted some of the major challenges the Internet faces, including repression of freedom of speech, denial of human rights, the spread of disinformation, the rise of cybercrimes, and balkanization.[284] In response, the Declaration outlines a vision which intends "to ensure that the use of digital technologies reinforces, not weakens, democracy and respect for human rights; offers opportunities for innovation in the digital ecosystem, including businesses large and small; and, maintains connections between our societies."[285] The Declaration also details a list of principles—including protection of human rights online, creating inclusive and affordable access to the internet, promoting trust in digital ecosystems, and protecting the 'multistakeholder internet governance.'[286] It is still too early to predict the effect or specific policies from this declaration, but it at least seems to suggest that the US government is starting to apply this form of soft power.

2. Inverse regulation: "Voluntary" Enforcement

Setting voluntary codes of conduct and guidelines can only go so far. To have an effect, enforcement is necessary. Consider TikTok's policy to confront PHM about COVID-19 vaccination. According to the policy, the platform would identify videos that use "words or hashtags related to the COVID-19 vaccine" and attach to those a banner that "redirects the user to verifiable, authoritative sources of information."[287] Researchers found that this policy was not enforced: 58% of relevant videos did not feature the banner.[288] Indeed, in platforms (as in governments), "law in the books and law in action" can be very different.[289]

---

[282] *Declaration for the Future of the Internet*, Bureau of Cyberspace and Digit. Pol'y, Dep't of State, https://www.state.gov/declaration-for-the-future-of-the-internet/ [https://perma.cc/QF6S-8LRG].

[283] Declaration for the Future of the Internet, https://www.state.gov/wp-content/uploads/2022/04/Declaration-for-the-Future-for-the-Internet.pdf [https://perma.cc/L7EL-5F48].

[284] *Id.*

[285] *Id.*

[286] *Id.*

[287] Kevin Morgan, *Taking action against COVID-19 vaccine misinformation*, TikTok – Community (Dec. 15, 2020), https://newsroom.tiktok.com/en-gb/taking-action-against-COVID-19-vaccine-misinformation [https://perma.cc/4DVF-MEQK].

[288] Ciarán O'Connor, *Tags, Flags & Banners: Evaluating the Application of Information Resources on Vaccine Content on TikTok*, Inst. for Strategic Dialogue—Digit. Dispatches (Nov. 4, 2021), https://www.isdglobal.org/digital_dispatches/tags-flags-and-banners-evaluating-the-application-of-information-resources-on-vaccine-content-on-tiktok/ [https://perma.cc/WKT9-JCGK].

[289] *See* Roscoe Pound, *Law in Books and Law in Action*, 44 Am. L. Rev. 12, 15 (1910).

Soft regulation in the form of government-platform cooperation can also feature in the *application* or *enforcement* of content moderation. Theoretically, states can require platforms to remove specific content using court orders. However, the slow and specific nature of the legal procedure is especially inapt for moderating fast-paced and high-volume online speech and PHM.[290] Instead, governments around the world opt for a more direct method to get platforms to act. Government actors identify content that they want removed from a specific platform and submit removal requests to platforms. Importantly, in this scheme, the governments' requests rely on the platforms' own terms of service, and the final decision regarding the removal of content remains with the platforms. This governmental use of platforms' private ordering is often referred to as "voluntary enforcement" or "inverse regulation."[291]

Major actors in the existing voluntary enforcement mechanisms are the Internet Referral Units ("IRU").[292] IRUs are the government actors that identify (themselves or with help from other governmental actors) content that should be removed, evaluate whether this content violates the platforms' terms of service, and issue take-down requests directly to the platforms.[293] As repeat players, IRUs are well versed in the platforms' internal governance, and they often enjoy a "trusted flagger" standing, which "prioritize[s]" their requests over others' requests.[294] While so far IRUs have

---

[290] Daphne Keller, *When Platforms Do the State's Bidding, Who Is Accountable? Not the Government, Says Israel's Supreme Court*, LAWFARE (Feb. 7, 2022), https://www.lawfareblog.com/when-platforms-do-states-bidding-who-accountable-not-government-says-israels-supreme-court [https://perma.cc/38G6-2AD8] ("If the goal is high-volume, high-speed resolution of speech claims, then the perfect (meaning judicial supervision) arguably becomes the enemy of the good (meaning public accountability of any sort).").

[291] *See, e.g.*, HCJ 7846/19 Adalah Legal Ctr. for Arab Minority Rts. in Isr. v. State Att'y's Off.—Cyber Dep't ¶¶ 6–7, 10, 45–50 (2021) (Isr.), *translated in* VERSA, Opinions of the Supreme Court of Israel, a Project of Cardozo Law, https://versa.cardozo.yu.edu/opinions/adalah-legal-center-arab-minority-rights-israel-v-state-attorney%E2%80%99s-office-%E2%80%93-cyber [https://perma.cc/J88V-XBG7]; Tomer Shadmy & Yuval Shany, *Protection Gaps in Public Law Governing Cyberspace: Israel's High Court's Decision on Government-Initiated Takedown Requests*, LAWFARE (Apr. 23, 2021), https://www.lawfareblog.com/protection-gaps-public-law-governing-cyberspace-israels-high-courts-decision-government-initiated [https://perma.cc/W25X-CS2E].

[292] *See generally* Rabea Eghbariah & Amre Metwally, *Informal Governance: Internet Referral Units and the Rise of State Interpretation of Terms of Service*, 23 YALE J.L. & TECH. 542 (2021); Brian Chang, *From Internet Referral Units to International Agreements; Censorship of the Internet by the UK and EU*, 49 COLUM. HUM. RTS. L. REV. 114 (2017).

[293] *See* Eghbariah & Metwally, *supra* note 292, at 556–57.

[294] *See* Eghbariah & Metwally, *supra* note 292, at 557, 563–64 (quoting Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y REV. 95, 108 (1974) (citing YouTube, *YouTube Trusted Flagger Program*, YOUTUBE HELP CTR., https://support.google.com/youtube/answer/7554338?hl=EN. [https://perma.cc/36FH-UKBN]) (last visited Mar. 3, 2023)). On trusted flaggers programs, see generally Naomi Appelman & Paddy Leerssen, *On "Trusted" Flaggers* (July 12, 2022), *in* PLATFORM GOVERNANCE TERMINOLOGIES ESSAY SERIES, https://law.yale.edu/sites/default/files/area/center/isp/documents/trustedflaggers_ispessayseries_2022.pdf [https://perma.cc/RY8X-N6BB].

mostly focused on terrorism and hate speech,[295] there are indications that IRUs have been flagging and filing removal requests about online PHM during the COVID-19 pandemic.[296] IRUs have been extremely successful in getting platforms to remove content, and they now operate in the UK, Israel, the EU, and various states in Europe.[297]

IRUs operations are seldom transparent,[298] but a recent case about Israel's IRU—the only legal case about IRUs to date—sheds some light on their operation. According to the Unit's internal procedure, revealed in the case, Israel's IRU will only reach out to platforms if all following conditions are met: (1) the content violates Israeli law; (2) the content violates a platform's standards; (3) the "severity" of the violation, "potential" spread, timeliness, or expected outcomes justify reaching out; (4) and balancing the constitutional rights of freedom of expression, "access to information," "privacy," human "dignity," "reputation," and "the public interest"—"justifies" notifying the platform.[299]

To the best of our knowledge, there is no IRU in the United States.[300] Establishing such a unit, or otherwise aggregating and systematizing the government's efforts to influence application of a platform's policies, is advisable. Indeed, we think the government should use voluntary enforcement to guide platform decisions about which content is undesirable and should be removed—for instance, PHM. However, some may disagree.

Opponents of voluntary enforcement would note that this mechanism allows governments to influence platforms' interpretation of their own policies.[301] In our view, increasing government involvement in application of rules that govern online speech is a blessing, not a curse. Of course, no one should underestimate the problems of giving the government too much power over speech. But this is a challenge that free speech doctrines in modern democracies understand and know how to confront. The alternative ap-

---

[295] *See, e.g.*, the EU's IRU most recent report, where the vast majority of content flagging was related to terrorism or violent extremism. *2020 EU IRU Transparency Report*, at 7, EU Internet Referral Unit (2021), https://www.europol.europa.eu/cms/sites/default/files/documents/EU_IRU_Transparency_Report_2020_2.pdf [https://perma.cc/FS56-KJY5].

[296] *See, e.g.*, *Enforcement Actions to Remove Content from Social Media Platforms 2020*, Israeli Internet Ass'n (Apr. 3, 2022) [hereinafter Israeli Internet Ass'n report], https://www.isoc.org.il/sts-data/cyber_unit_2020 [https://perma.cc/Z4KU-U9C2] (noting that in 2020 2.3% of requests originated from health officials with regard to public health misinformation).

[297] *See* Eghbariah & Metwally, *Informal Governance*, *supra* note 292, at 567–86 (discussing establishment and actions of IRUs in UK, EU, Israel, France, and the US); Chang, *supra* note 292, at 126–43 (discussing the establishment and operations of the UK and EU IRUs).

[298] Eghbariah & Metwally, *Informal Governance*, *supra* note 292, at 564–66; Full Fact, Full Fact Report 2022: Tackling Online Misinformation in an Open Society—What Law and Regulation Should Do 62–67 (2022), https://fullfact.org/media/uploads/full-fact-report-2022.pdf [https://perma.cc/H9B2-3S23] (noting the lack of transparency over UK's IRU).

[299] HCJ 7846/19, *supra* note 291, at ¶¶ 7–11.

[300] *See* Eghbariah & Metwally, *Informal Governance*, *supra* note 292, at 583–85.

[301] *Id.*

proaches, leaving regulation of online speech only to platforms or relying on ordinary legal proceedings, are problematic.[302] Thus, it is doubtful that the government could achieve the goals of voluntary enforcement mechanisms with less restrictive means.[303] To the extent that governments should have some influence on the application and enforcement of online PHM policies, it must cooperate with the platforms. Voluntary enforcement is one mechanism that allows this influence.

Critics may raise doubts whether such mechanisms can actually be voluntary. According to this view, platforms do not enjoy unlimited discretion in responding to IRU's requests because governments exert powerful leverage over the platform.[304] So, the argument goes, government could coerce platforms by threatening to take legal action against them, whether or not the government currently has the authority to use such measures.[305] The Israeli Supreme Court made similar claims about voluntary enforcement. It held that the mere possibility that governments might devise compulsory regulation at any time hinders the voluntariness of the enforcement.[306] This criticism finds legal grounding in *Bantam Books v. Sullivan*.[307] In that case, a government commission sent letters to booksellers, indicating the commission found some of their books objectionable, notifying the sellers that the commission had contacted the attorney general and police, and thanking the booksellers in advance for their cooperation. Justice Brennan, writing for the Court, held that while the commission did not censure or prosecute the booksellers, it tried to censor and suppress the publication. This scheme, Brennan held, amounted to an administrative prior restraint, which bears "a heavy presumption against its constitutional validity."[308]

This criticism does not carry the same weight in the context of online PHM. As Justice Brennan emphasized, crucial to the decision that the state censored the booksellers in *Bantam Books* was the factual finding that the booksellers' compliance was not voluntary, but caused by governmental in-

---

[302] *See supra* Part II and Keller, *supra* note 290, respectively.

[303] In addition, using this mechanism to confront PHM seems to serve a compelling state interest. *See, e.g.*, Does 1-6 v. Mills, 16 F.4th 20, 32 (1st Cir. 2021), *cert. denied sub nom.* Does 1-3 v. Mills, 142 S. Ct. 17 (2021) (noting that "[s]temming the spread of COVID-19 is . . . a compelling interest" (quoting Roman Cath. Diocese of Brooklyn v. Cuomo, 141 S. Ct. 63, 67 (2020)).

[304] *See supra* text accompanying notes 248–54; Keller, *supra* note (("For a strictly rational platform, saying no to governments may not be worth the potential costs.").

[305] *See* Bambauer, *Against Jawboning*, *supra* note 150, at 55.

[306] HCJ 7846/19, *supra* note 291, at ¶ 51.

[307] 372 U.S. 58 (1963); *see also* Genevieve Lakier, *Informal Government Coercion and The Problem of "Jawboning,"* LAWFARE (July 26, 2021), https://www.lawfareblog.com/informal-government-coercion-and-problem-jawboning [https://perma.cc/JS8C-X3F7].

[308] *Bantam Books*, 372 U.S. at 59–71 ("The effect of the [letters] were (sic) clearly to intimidate the various book and magazine wholesale distributors and retailers and to cause them, by reason of such intimidation and threat of prosecution, (a) to refuse to take new orders for the proscribed publications, (b) to cease selling any of the copies on hand, (c) to withdraw from retailers all unsold copies, and (d) to return all unsold copies to the publishers.").

timidation.[309] Recently, in *Speech First, Inc. v. Cartwright*, the Eleventh Circuit reiterated this point.[310] It held that examining whether the intended audience of the government's communication was in fact, or would likely be, cowed by such communication, is crucial for application of *Bantam Books*.[311]

Arguably, *Bantam Books* should not ordinarily apply to a voluntary enforcement mechanism. For one, the voluntary enforcement mechanism relies on platforms' agreement to apply private ordering that the platforms themselves adopted. Both aspects—that the compliance to the state's requests is voluntary, and that the platforms set these norms, not the state—distinguish this mechanism from *Bantam Books*. And in practice, platforms do push back against such requests. In the case of voluntary enforcement by platforms, most recent data suggests that platforms reject more than 25% of requests from IRUs, meaning that they are in fact free to reject states' requests.[312] This conclusion makes sense given the titanic nature of platforms, as international companies with unimaginable financial resources, and the extensive protection that § 230 provides.[313] Platforms' transparency about the government's efforts to pressure them are another valuable tool platforms can use to push back at the government's requests.[314] Against this background, the claim that platforms are weak and easily suppressed is not convincing.[315] Thus, it is doubtful that IRUs' actions, or other means of voluntary enforcement, involve coercion or prior restraint. Therefore, they are likely permissible under *Bantam Books*.

Indeed, recent cases have refused to extend *Bantam Books* to platforms. In two separate cases, Twitter argued that public expressions of support by government officials for regulating the company are analogous to the Commission's letters in *Bantam*, and thus constitute prior restraint and censorship.[316] The courts rejected these claims and refused to apply the *Bantam Books* factors.[317] Additionally, in *VDARE Foundation v. City of Colorado Springs*, the Tenth Circuit held that government may communicate to private actors without running afoul of *Bantam Books*, so long as those communications do not include threats of legal action or prosecution.[318] The court found

---

[309] *Id.* at 63–64, 68.

[310] 32 F.4th 1110 (11th Cir. 2022).

[311] *Id.* at 1122–24.

[312] Israeli Internet Ass'n report, *supra* note 296 (noting that overall, 27% of Israel's IRU requests were rejected by the Platforms, a sharp increase compared to ~10% in 2019); *6th Evaluation*, *supra* note 267 ("IT companies removed 62.5% of the content notified to them, while 37.5% remained online.").

[313] *See supra* Section IV.G.

[314] *See* Bambauer, *Against Jawboning*, *supra* note 150, at 111–13.

[315] *See id.* at 59–60, 85–87.

[316] *See* Twitter, Inc. v. Paxton, 26 F.4th 1119 (9th Cir. 2022); Trump v. Twitter Inc., 602 F. Supp. 3d 1213 (N.D. Cal. 2022).

[317] *See Paxton*, 26 F.4th at 1126–27; *Trump*, 602 F. Supp. 3d at 1220–24 (N.D. Cal. 2022).

[318] VDARE Found. v. City of Colorado Springs, 11 F.4th 1151, 1165–73 (10th Cir. 2021), *cert. denied*, 142 S. Ct. 1208 (2022).

that the government's communications are merely permissible government speech.[319] Similarly, in *Kennedy v. Warren*, a Senator wrote a letter to Amazon asking the company to amend its algorithms so that they would not promote books that propagate COVID-19 PHM.[320] The court distinguished this case from *Bantam Books*, holding that the letter exhibited no regulatory power, no threat of enforcement, and no "realistic chance the threatened action c[ould] be carried out."[321] Recent case law suggests that the voluntary enforcement mechanism does not undermine the voluntariness of platforms' actions.

A final criticism against adopting the voluntary enforcement mechanism focuses on the speaker's perspective. Individuals whose speech was limited by virtue of such voluntary enforcement seldom have a good recourse; they seldom know that the state was involved in removing their content and can, at best, try and plead with the platforms.[322] This is troubling given the ongoing expansion of topics on which governments use voluntary enforcement. What began with hate speech, child pornography, and terrorism was recently supplemented with confronting online PHM and more recently "urgent action to limit disinformation related to the war in Ukraine."[323] In this Article we support the need for actions against online PHM based on its specific nature and the harms it poses, particularly in times of pandemics. The use of voluntary enforcement for other topics requires separate analysis and scrutiny which is beyond our scope. But critics of voluntary enforcement are correct to warn against excessive or coercive use of this method.[324] We agree that the voluntary enforcement mechanism, like other government actions, should be backed by a legitimate authorization process, clear procedures, and rules, be open to judicial review and other public scrutiny, and be as transparent as possible. Existing IRUs illustrate that voluntary enforcement mechanisms can satisfy those characteristics—they can be tamed and scrutinized by judicial review and other checks and balances.

B. Reform Proposals: Regulating Algorithmic Amplification

As discussed earlier, governments can disseminate reliable public health information and influence the behavior of social media platforms using soft law techniques. They should do more to take advantage of these techniques. However, soft law leaves much to the discretion of platforms, which may choose to ignore government requests or push back if these re-

---

[319] *Id.*

[320] Kennedy v. Warren, No. 2:21-CV-01508-BJR, 2022 WL 1449678, at *5 (W.D. Wash. May 9, 2022).

[321] *Id.* at *5.

[322] *See generally* Eghbariah & Metwally, *Informal Governance*, *supra* note 292.

[323] *2022 Code of Practice on Disinformation*, *supra* note 268.

[324] Bambauer, *Against Jawboning*, *supra* note 150, at 87–92.

quests become too onerous. Hence, mandatory measures have considerable appeal. The trouble is that the First Amendment seems to block regulations of online misinformation.[325] In this Section, we argue that a particularly influential aspect of online PHM—amplification of content via algorithmic recommendation—can be regulated. In other words, First Amendment doctrine poses fewer obstacles than commonly supposed.

Why regulate recommendation algorithms? Renee DiResta's insight that "free speech is not the same as free reach"[326] is illuminating. As DiResta observes, "There is no right to algorithmic amplification. In fact, that's the very problem that needs fixing."[327] And it needs fixing because platforms rely on surreptitious data collection and profiling practices to optimize highly engaging content on a personalized basis.[328] There is strong evidence that ordinary factual content is not very engaging as compared with misinformation[329] or various forms of abusive, divisive, polarizing, and extremist content.[330] These outcomes can be attributed to platforms' failure to address the harmful and discriminatory consequences of ranking content on a personalized basis and also to platforms' decision to reward "borderline" content with an algorithmic boost.[331]

---

[325] See *supra* Section IV.F.

[326] Renee DiResta, *Free Speech Is Not the Same As Free Reach*, WIRED (Aug. 30, 2018), https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/ [https://perma.cc/PCD8-EEVK].

[327] *Id. See also* Erin L. Miller, *Amplified Speech*, 43 CARDOZO L. REV. 1, 5 (2021) ("As speech reaches larger and larger audiences, it has a *smaller* impact on the speaker's own interests, properly understood, and has a *greater* impact on democratic discourse. . . . But past some threshold audience size, just *adding* listeners does little to enhance these characteristics, and can actually undermine them."); *cf.* Daphne Keller, *Amplification and Its Discontents: Why Regulating the Reach of Online Content Is Hard*, 1 J. FREE SPEECH L. 227 (2021).

[328] That is, content that best captures user attention and most increases time spent on the platform (and hence advertising revenues). *See generally* SINAN ARAL, THE HYPE MACHINE: HOW SOCIAL MEDIA DISRUPTS OUR ELECTIONS, OUR ECONOMY, AND OUR HEALTH—AND HOW WE MUST ADAPT 200–25 (2020) (describing how online platforms use indirect data to target content toward specific users); TAINA BUCHER, IF . . . THEN: ALGORITHMIC POWER AND POLITICS 5–6 (2018) (explaining how Facebook communicates about "friends" to users).

[329] *See* Vosoughi et al., *supra* note 40, at 1146.

[330] *See, e.g.*, Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (May 5, 2021), https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/ [https://perma.cc/298J-KHHL] (Facebook's research indicates that "no matter where we draw the lines for what is allowed, as a piece of content gets close to that line, people will engage with it more on average."); EU Counter-Terrorism Coordinator, *The Role of Algorithmic Amplification in Promoting Violent and Extremist Content and Its Dissemination on Platforms and Social Media* (Dec. 9, 2020), https://data.consilium.europa.eu/doc/document/ST-12735-2020-INIT/en/pdf [https://perma.cc/7RD4-WW96].

[331] *See* Keach Hagey & Jeff Horwitz, *Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead.*, WALL ST. J. (Sept. 15, 2021), https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215 [https://perma.cc/P92H-46RZ] (describing how Facebook's leadership rejected suggestions to make its algorithms less rewarding towards outrage and lies because the change could undermine user engagement); Karen Hao, *How Facebook Got Addicted to Spreading Misinformation*, MIT TECH. REV. (Mar. 11, 2021), https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/ [https://perma.cc/8ZMC-ZSBR] (explaining that Facebook rejected proposals

Congress is considering this reform path. Recent bills seek to limit these negative impacts by better regulating data collection and profiling,[332] forcing platforms to disclose their use of such data for algorithmic ranking and amplification purposes,[333] taking better account of the harms associated with algorithmic ranking systems notwithstanding Section 230 immunity,[334] and even prohibiting the use of such systems entirely to the extent that they lead to civil rights violations.[335] Other proposals from academics and think tanks include "middleware" solutions that would outsource algorithmic ranking systems to third parties, thereby giving users more control over their online experiences while "prevent[ing]" dominant platforms 'from using their power to artificially amplify or suppress certain types of speech' "[336] and content-neutral "friction" measures such as communication delays that reduce the velocity of network sharing, virality "speed bumps" that restrict the scale and scope of viral sharing, and various transparency measures.[337]

All of these bills and proposals share a common goal of mitigating the harms associated with the use of algorithmic ranking systems that optimize engagement to maximize corporate revenues. Although they differ in various ways, all of them must comport with the First Amendment, which imposes heavy burdens on regulating both content moderation and algorithmic ranking. A bedrock of First Amendment doctrine is that a private person or entity

---

to change amplification algorithms that would reduce political polarization). *But see* Keller, *Amplification and Its Discontents*, *supra* note 327, at 230 n.3 (pointing to doubts about the role of amplification algorithms in the increased popularity of some extremist content).

[332] *See* Consumer Online Privacy Rights Act, S. 3195, 117th Cong. § 108(a)(1) (2021) (prohibiting entities from "process[ing] or transfer[ring] . . . data on the basis of" specified protected characteristics like race, religion, and gender); Online Privacy Act of 2019, H.R. 4978, 116th Cong. § 106 (2019) (requiring users to opt-in before a platform can process their personal data using an algorithm for purposes of "behavioral personalization").

[333] *See* Filter Bubble Transparency Act, S. 2024, 117th Cong. (2021) (requiring platforms using an "algorithmic ranking system" to (1) notify users that they use their data to curate their experiences and (2) allow users to opt-out of this version of the service in favor of an "algorithm-free" version); Algorithmic Justice and Online Platform Transparency Act, S. 1896, 117th Cong. (2021) (requiring platforms to explain to users what kinds of personal information they collect to enable algorithmic processes, how they collect this data, how they use this data to train or facilitate algorithmic processes, and how these algorithmic processes use this data to curate user's experiences).

[334] *See generally* Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117th Cong. (2021) (amending Section 230 to remove liability protection from platforms that use algorithms to rank the delivery of information, unless they sort information in specified ways deemed less harmful than algorithmic ranking).

[335] *See* S. 1896 (prohibiting algorithmic processes on online platforms that discriminate on the basis of race, age, gender, ability, and other protected characteristics and establishing a safety and effectiveness standard for algorithms, such that online platforms may not employ automated processes that harm users or fail to take reasonable steps to ensure algorithms achieve their intended purposes).

[336] *See* Francis Fukuyama, *Making the Internet Safe for Democracy*, 32 J. DEMOCRACY 37, 40, 41–43, 44 (2021).

[337] Ellen P. Goodman, *Digital Fidelity and Friction*, 21 NEV. L.J. 623, 646 (2021); *see also* Brett M. Frischmann & Susan Benesch, Friction-in-Design Regulation as 21st Century TPM (Aug. 1, 2022) (unpublished manuscript), https://ssrn.com/abstract=4178647 [https://perma.cc/7CU7-FPFE].

is protected against government efforts to prescribe what they may say, how they say it, or to compel them to speak contrary to their will.[338] Subject to limited exceptions (like fraud, obscenity, incitement to imminent violence, speech integral to criminal conduct, and so on), the First Amendment "demands that content-based restrictions on speech be presumed invalid . . . and that the Government bear the burden of showing their constitutionality" under the appropriate constitutional standard.[339]

This Section argues that proposals seeking to regulate algorithmic amplification would not violate the First Amendment. Rather, they are content-neutral restrictions that would not interfere with the rights of social media platforms to moderate content as they see fit.[340] Treating such laws as if they run afoul of the First Amendment rests on a conceptual confusion. Namely, it conflates content-moderation—which is inherently a content-based activity subject to the highest level of First Amendment scrutiny—with amplification or ranking—which is a content-neutral task subject to a lower level of scrutiny. Correcting this confusion is important, because a well-drafted law restricting amplification may indeed survive intermediate scrutiny.[341]

We elaborate on these points by analyzing the differences between content moderation and algorithmic ranking (which we also refer to as recommendation or amplification). Next, we argue that these technological distinctions merit different First Amendment analysis. We show how this distinction plays out in the recent controversy over state "anti-censorship" laws leading to conflicting lower court decisions and a possible showdown in the Supreme Court.[342] And we show why legislative proposals seeking to regulate algorithmic amplification may survive First Amendment scrutiny if properly analyzed. Finally, we briefly review the benefits of content-neutral regulation of algorithmic ranking in addressing the harms of misinformation (including online PHM).

*****

Before embarking on this discussion, however, it is important to explain why we are not considering any proposals limited solely to PHM. The reason is simple: any law aimed at reducing amplification of any specified subject matter (apart from the usual exceptions like obscenity) would be treated as a content-based restriction, reviewed under strict scrutiny, and

---

[338] *See, e.g.*, West Virginia State Board of Education v. Barnette, 319 U.S. 624, 642 (1943) (underscoring the constitutional protection against compelled speech).

[339] United States v. Alvarez, 567 U.S. 709, 716–17 (2012) (quoting Ashcroft v. American Civil Liberties Union, 542 U.S. 656, 660 (2004)).

[340] *Cf.* Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 256 (1974) (finding regulations that required newspapers to publish content interfered with publishers' speech rights).

[341] *See* Miller, *Amplified Speech*, *supra* note 327, at 15.

[342] *See* Robert Barnes & Ann E. Marimow, *A Landmark Supreme Court Fight over Social Media Now Looks Likely*, WASH. POST (Sept. 19, 2022), https://www.washingtonpost.com/politics/2022/09/19/texas-florida-social-media-laws/ [https://perma.cc/M9FL-8WSK].

likely struck down. The most that any legislative reforms can hope to achieve, consistent with the First Amendment, are the sort of across-the-board restrictions mentioned above: limits on the use of personal data for targeted recommendations, better disclosure and accountability measures, or content-neutral restrictions on amplification. We believe that several of these approaches may survive First Amendment scrutiny and help reduce the harms associated with PHM (and other forms of misinformation too). Our goal, therefore, is to clear away First Amendment obstacles that would impede regulatory responses.[343] Under existing First Amendment doctrine, this is the best way to try and tame online PHM.

1. Content Moderation vs. Algorithmic Ranking

A defining feature of successful social media platforms like Facebook, YouTube, and Twitter is an abundance of user-generated content. The sheer number of social media users and the vast scale and complexity of what they post (Facebook users alone share approximately 4.75 billion items each day) forces large platforms to rely on automation to manage this content and satisfy the desire of users to see what they consider relevant.[344] The two main procedures they apply are content moderation and algorithmic ranking.

How does automated content moderation work? A recent technical primer defines algorithmic content moderation as a system that "classif[ies] user-generated content based on either matching or prediction, leading to a decision and governance outcome (e.g., removal, geoblocking, account takedown)."[345] The aim of matching is detection, that is, determining whether a given string of data and some other string of data refer to the same text, audio, image or video.[346] Matching's counterpart, classification (which is a form of prediction), involves assessing "newly uploaded content that has no corresponding previous version in a database; . . . the aim is to put new content into one of a number of categories."[347]

To successfully classify uploaded content, platforms like Facebook rely on machine learning techniques. Generally, the process involves two main steps: (1) human review and evaluation of sample texts to determine what

---

[343] *See generally* Frischmann & Benesch, *supra* note 337 (describing how the First Amendment is a strong barrier to regulation of online content).

[344] *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 97–110 (2018).

[345] Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, 7 BIG DATA & SOC'Y 1, 3 (2020).

[346] Matching also typically involves a technique known as hashing, which is the process of "transforming a known example of a piece of content into a 'hash' – a string of data meant to uniquely identify the underlying content" (such as known images of child pornography). *Id.* at 4.

[347] *Id.* at 5; *see also* NOAH GIANSIRACUSA, HOW ALGORITHMS CREATE AND PREVENT FAKE NEWS: EXPLORING THE IMPACTS OF SOCIAL MEDIA, DEEPFAKES, GPT-3, AND MORE 202 (2021).

category they belong in (e.g., "hate speech" or "threats of violence"); and (2) training classifiers to predict whether some unknown texts fit into these categories.[348] These systems are imperfect, producing both false positives and false negatives.[349] They also require both automated and human review to handle any sensitive content, requiring a more nuanced and contextual approach than algorithmic systems alone can achieve at present.[350] For large social media platforms, content moderation requires enormous resources.[351] One thing is certain about content moderation: whatever techniques it may rely on to match or classify content, from a First Amendment perspective, it necessarily targets speech based on its message or subject matter and is the very paradigm of a content-based activity.

In contrast, algorithmic ranking is "designed to estimate the utility of an item and predict whether it is worth recommending."[352] Ranking or amplifying content also relies on machine learning algorithms to overcome problems of number and scale.[353] This task differs from content moderation in both its methods and goals. Algorithmic ranking is largely indifferent to content. The category a piece of content belongs in may count as one of hundreds of relevant factors, but the goal of a ranking algorithm is to meet a defined utility function—for example, most video clicks or views, longest watch time, or greatest user satisfaction—and then devise an algorithm that is optimized to achieve this value.[354] Importantly, it makes no difference to the design of such algorithms whether a user prefers cat videos, gaming videos, or PHM videos denouncing the CDC and the Gates Foundation. The algorithm is successful if it correctly predicts that users will remain engaged with relevant content, spend more time on the platform, view more ads, and purchase more advertised products and services, thereby increasing platform revenues and profits. In short, ranking algorithms are engagement-driven but content-neutral.

A deeper dive into Facebook's newsfeed algorithm helps drive home the differences between content moderation and algorithmic ranking. Facebook's

---

[348] *See* Gorwa et al., *supra* note 345, at 5.

[349] *See* GILLESPIE, *supra* note 344, at 104–05.

[350] *See id.* at 104–07.

[351] Facebook has almost 40,000 people working on safety and security issues. *See* Kurt Wagner, *Facebook Says It Has Spent $13 Billion on Safety and Security Efforts Since 2016*, FORTUNE (Sept. 21, 2021), https://fortune.com/2021/09/21/facebook-says-it-has-spent-13-billion-on-safety-and-security-efforts-since-2016/ [https://perma.cc/QW2J-HHE9].

[352] Qian Zhang, Jie Lu & Yaochu Jin, *Artificial Intelligence in Recommender Systems*, 7 COMPLEX & INTELLIGENT SYS. 439, 440 (2021).

[353] *See* Akos Lada, Meihong Wang & Tak Yan, *How Machine Learning Powers Facebook's News Feed Ranking Algorithm*, ENGINEERING AT META (Jan. 26, 2021), https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/ [https://perma.cc/93LC-DUUB] ("We need to score all the posts available for more than 2 billion people (more than 1,000 posts per user, per day, on average), which is challenging. And we need to do this in real time.").

[354] *See* Paige Cooper, *How the YouTube Algorithm Works in 2023: The Complete Guide*, HOOTSUITE (June 21, 2021), https://blog.hootsuite.com/how-the-youtube-algorithm-works/ [https://perma.cc/W6KQ-WHCE].

newsfeed ranks content according to its unique relevance to specific users based on predictive models that learn what drives us to interact with a piece of content. According to Sinan Aral, "The models predict whether we will engage with the content based on who posted it, what's it about, whether it contains an image, or a video, what's in the video, how recent it is, how many of our friends liked or shared it and so on."[355] Taina Bucher offers a similar explanation, noting that Facebook's newsfeed algorithm assigns a "relevancy score" to specific Facebook users based on user activity such as "friend relationships, frequency of interactions, number of likes and shares a post receives, how much a user has interacted with particular types of posts in the past" and so on, and then sorts posts into the preferred order for each user based on these relevancy scores.[356] Thus, content is just one of many signals in Facebook's newsfeed algorithm—and seemingly a much weaker signal than numerous non-content signals.[357] The same is probably true for YouTube and other platforms.[358] In any event, the meaning or subject matter of the content is not what drives its ranking or amplification on the newsfeed algorithm.[359]

Additionally, it is important to keep in mind the scale at which platforms' ranking system operates. According to Facebook, the system "need[s] to score all the posts available for more than two billion people (more than 1,000 posts per user, per day, on average) . . . in real time."[360] This is enormously challenging both computationally and algorithmically. Facebook relies on what it calls a "feed aggregator" to "collect all relevant information about a post and analyze all the features . . . in order to predict

---

[355] ARAL, *supra* note 328, at 84 (noting that Facebook's algorithm considers about two thousand factors).

[356] BUCHER, *supra* note 328, at 78.

[357] Indeed, a Facebook blog post currently describing "How Feed Works" identifies the "three main signals" used to estimate relevance as (1) "Who posted it" (i.e., friends, family, news sources, businesses, public figures, etc.); (2) "Type of content" (i.e., photos, videos, or links); and (3) "Interactions with the post" (i.e., likes, reactions, comments, and shares). *See Feed Shows You Stories That Are Meaningful and Informative*, META, https://www.facebook.com/formedia/tools/feed [https://perma.cc/4CVH-PZUP].

[358] *See* YOUTUBE REGRETS, MOZILLA FOUND. 13–19 (2021), https://foundation.mozilla.org/en/youtube/findings/ [https://perma.cc/7H5M-BXCM].

[359] For the sake of clarity, the newsfeed algorithm should not be confused with advertising mechanisms like voter microtargeting. The goal of the newsfeed algorithm is to decide which of hundreds of posts to amplify based on the personal characteristics and activities of a given user. In voter microtargeting, on the other hand, a political campaign uses predictive modeling of a voters' individual preferences to target specific messages at specific people. The goal is "[p]ersonalized [m]ass [p]ersuasion," hence the model determines which voters to target and how to adjust the message "to maximize reach, awareness, and influence among the voters . . . it is trying to persuade." ARAL, *supra* note 328, at 131, 145–48, 206. For example, the campaign builds a model of voters likely to support a pro-gun control message based on their demographics, behaviors, preferences, social networks, and location histories. *See id.* at 146, 205. It then adjusts the message depending on a range of factors. *See id.* at 212–15. Thus, voter microtargeting, unlike the ranking of newsfeed posts, is inherently a content-based activity because it always begins with a specific message, whereas the newsfeed algorithm ranks messages (whatever their content happens to be) according to their relevance to a given user.

[360] Lada et al., *supra* note 353.

the post's value . . . to the user, as well as the final ranking score . . . by aggregating all the predictions" from multiple prediction models.[361] In technical terms, this means analyzing up to a thousand signals using multitask "neural nets" that after several "passes" over the eligible content spits out a relevancy score resulting in a personalized newsfeed for every Facebook user.[362] For our purposes, mastering the technical details is not crucial. What's crucial is understanding that algorithmic ranking is not (or at least only marginally) sensitive to specific content. Hence, use of recommendation algorithms is ordinarily content neutral.

So, the backend of ranking systems consists in an elaborate scheme of algorithmic ordering and evaluation. This scheme analyzes copious amounts of data, far more than humanly possible, based on far more factors than humans can navigate. The scale of this system, alongside its technological features as indifferent to content, almost compels it to be content-neutral. That is, the use of recommendation algorithms is *not* an expressive activity, it does not express platform's speech. Thus, the use of ranking algorithms is readily distinguishable from the exercise of editorial judgement by human editors as they select and thoughtfully organize the content of a newspaper (a point we return to below).

Several courts have recognized the content-neutrality of recommendation algorithms, in the context of deciding whether their use results in a loss of Section 230 immunity. For example, in *Dyroff v. Ultimate Software Group, Inc.*, the Ninth Circuit described data mining and recommendation algorithms as "content-neutral" tools for Section 230 purposes.[363] Similarly, in *Force v. Facebook, Inc.*, the Second Circuit described Facebook's algorithms as content-neutral insofar as they "take the information provided by Facebook users and 'match' it to other users . . . based on objective factors applicable to any content, whether it concerns soccer, Picasso, or plumbers."[364] The decision in *NetChoice v. Attorney General* more explicitly sheds light on the constitutionality of regulating algorithmic amplification.[365] This case addressed a Florida law known as S.B. 7072,[366] Florida's so-called "anti-censorship" law. This law restricts the ability of social media firms to engage in content moderation.[367] It also requires platforms to allow users to opt out on annual basis of "post-prioritization" (promoting or demoting content in a newsfeed) or "shadow banning" (removing or reducing the visibility of a user's content without telling them) and instead receive content in "sequential or chronological" order.[368] According to the Eleventh Circuit,

---

[361] *Id.*

[362] *Id.* For a good overview of neural networks intended for a non-technical audience, see PANOS LOURIDAS, ALGORITHMS 181–230 (2020).

[363] 934 F.3d 1093, 1096 (9th Cir. 2019), *cert. denied*, 140 S. Ct. 2761 (2020).

[364] 934 F.3d 53, 70 (2d Cir. 2019).

[365] 34 F.4th 1196 (11th Cir. 2022).

[366] FLA. STAT. §§ 106.072, 501.2041.

[367] *Id.*

[368] FLA. STAT. §§ 501.2041(1)(e)–(f), (2)(f) (2022).

this opt-out provision is "pretty obviously content-neutral" because "a requirement that platforms allow users to decline content curation" does not "depend[ ] in any way on the substance of the platforms' content-moderation decisions."[369]

In sum, content-moderation and algorithmic recommendation are technologically and analytically distinct. This distinction matters for First Amendment analysis. Regulating the former is ordinarily content-based and calls for strict scrutiny analysis. Conversely, regulating the latter is ordinarily content-neutral, and thus faces only the hurdle of intermediate scrutiny. Hence, laws that try to regulate online PHM through content ranking—i.e., regulation of algorithmic recommendation or amplification—face only the more lenient standard. Previous analysis in this Article explained that confronting online PHM is a compelling state interest.[370] Therefore, properly drafted laws of this kind are well situated to withstand First Amendment challenges.

2. *Applying the Distinction to Protected Editorial Judgment*

One major point of contention about regulation of social media platforms pertains to their editorial judgment. "[E]ditorial . . . judgment," the Supreme Court explained in *Miami Herald Publishing Co. v. Tornillo*,[371] is "[t]he choice of material to go into a newspaper, and the decisions made as to the limitations on the size and content of the paper, and treatment of public issues and public officials."[372] That case held that right-of-reply statutes are unconstitutional because they are an "intrusion into the function of editors."[373] For our purposes, the question is whether the use of recommendation algorithms to manage the content that users see amounts to an expressive editorial judgment. If this is the case, then this very common use of recommendation algorithms is protected by the First Amendment, and our calls to regulate them are hopeless. We think it is not the case. Recognizing the distinction between content moderation and recommendation algorithms shows why.

To explain this point, consider a recent circuit split regarding the constitutionality of Florida and Texas laws, respectively, regulating social media platforms. In *NetChoice v. Attorney General*, the Eleventh Circuit invali-

---

[369] NetChoice v. Att'y Gen., 34 F.4th 1196, 1226 (11th Cir. 2022).

[370] *See supra* Section II.B.

[371] 418 U.S. 241 (1974).

[372] *Id.* at 258. *See also* Pac. Gas & Electric Co. v. Pub. Utils. Comm'n of California, 475 U.S. 1, 10 (1986) (holding that utility company is not required to include in its billing envelopes a message with which it disagreed); Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 636–37 (1994) (noting that the decisions of cable operators about which channels to offer were protected speech); Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston, 515 U.S. 557, 570 (1995) (holding that "[t]he selection of contingents to make a parade is entitled to similar protection" to "the presentation of an edited compilation of speech generated by other persons").

[373] *Miami Herald Publ'g Co.*, 418 U.S. at 258.

dated the content-moderation provisions of the Florida law.[374] In doing so, the court extended First Amendment protection of "editorial judgment" from newspapers and other traditional media to social media platforms. The Eleventh Circuit's main reasoning was that platforms' content-management decisions are analogous to a newspaper's exercise of editorial discretion. "By engaging in content moderation," the court notes, "platforms develop particular market niches, foster different sorts of online communities, and promote various values and viewpoints."[375] The court therefore held that "platforms' . . . decisions" about ordering third-party content—including "whether, to what extent, and in what manner to disseminate" such "content to the public are editorial judgments protected by the First Amendment."[376] Conversely, in *NetChoice v. Paxton*, the Fifth Circuit analyzed a similar Texas law but reached an opposite conclusion about editorial judgment.[377] Notably, the Fifth Circuit held that "[u]nlike newspapers, . . . [p]latforms exercise virtually no editorial control or judgment" over the content shared on their services "and use sophisticated algorithms to arrange and present . . . it."[378] Consequently, they cannot claim that their content moderation decisions are expressive conduct protected by the First Amendment.

There are several reasons to think that platforms engage in editorial judgement when they moderate content. Every day large platforms make choices about the content they wish to host, demote, or remove.[379] They do so by devising principles and policies that both set expectations for users and justify how platforms will handle inevitable controversies over offensive speech. These rules take the form of community guidelines that "articulate the 'ethos' of a site, not only to lure and keep participants, but also to satisfy

---

[374] *NetChoice*, 34 F.4th at 1203. S.B. 7072 uses various measures to gut the ability of social media platforms to develop and apply platform-specific rules governing permissible speech on the platform. It blocks platforms from deplatforming political candidates or (again with respect to candidates during an election) engaging in "post-prioritization or shadow banning." FLA. STAT. § 501.2041(2)(h). The law also prohibits "any action to censor, deplatform, or shadow ban a journalistic enterprise based on . . . content"; requires platforms to enforce content-moderation standards "in a consistent manner"; and allows users to turn off algorithmic ranking in favor of sequential or reverse chronological ordering. FLA. STAT. §§ 501.2041(2)(b), (f)(2), (j).

[375] *NetChoice*, 34 F.4th at 1205.

[376] *Id.* at 1212.

[377] 49 F.4th 439 (5th Cir. 2022). H.B. 20 is codified at Texas Business and Commerce Code §§ 120.001–151 and Texas Civil Practice and Remedies Code §§ 143A.001–08. The law restricts content-moderation policies in a very straightforward fashion: Section 143A.002(a)(1), (3) of the Texas Civil Practice and Remedies Code states that "a social media platform may not censor a user, a user's expression, or a user's ability to receive the expression of another person based on . . . the viewpoint of the user or another person," or the "user's . . . location." Texas Civil Practice and Remedies Code §§ 143A.002(a)(1), (3). The law also requires social media platforms to disclose their content and data management procedures, produce regular reports of removed content, and create a "complaint system." Texas Business and Commerce Code §§ 120.051, 120.053, 120.101.

[378] *Paxton*, 49 F.4th at 440, 464.

[379] *See* GILLESPIE, *supra* note 344, at 21 (describing content moderation as essential, even constitutional, activity of platforms).

the platform's founders, managers, and employees, who want to believe that the platform is in keeping with their own aims and values."[380] Even though the purpose and form of such guidelines are fairly consistent across platforms, their content differs markedly depending on how a company understands its own mission relative to the user community it wishes to cultivate.[381] Established platforms like Facebook tend to agonize over the tension between its commitment to open expression and the need to limit expression to prevent abuse. But even the newer conservative platforms that promise their users a "censorship-free" experience eventually have to deal with offensive speech and bad actors and do so by developing community guidelines that express their own values and viewpoints.[382] In both cases, these guidelines are expressive—they express what a company stands for.

Thus, the Eleventh Circuit seems to have it right with regards to content moderation. But the court takes a wrong turn when it characterizes algorithmic amplification—in this case, post-prioritization and shadow banning—as inherently expressive.[383] This essentially means that the use of

---

[380] *Id.* at 47.

[381] A comparison of Facebook and the newer conservative social media platforms drives this point home. Facebook's Community Standards express a commitment to "giv[ing] people a voice." *Facebook Community Standards*, META, https://transparency.fb.com/policies/community-standards/ [https://perma.cc/ACR7-2FE6]. Facebook acknowledges that giving everyone a voice may lead to disagreement or even some content that users find offensive, but it promises to limit expression only in the service of four company values: authenticity, safety, privacy, and dignity. *See id.* The company's Community Standards encompass six main categories and twenty-four sub-categories, each of which lays out a policy rationale and specific, detailed prohibitions on posting materials related to violence and criminal behavior, safety threats, objectionable content, integrity and inauthenticity, and intellectual property violations. *See id.* Parler (like Rumble, Gettr, and Truth Social) has very different policies reflecting its own distinctive values and viewpoints, especially regarding hate speech and misinformation. Parler's Community Guidelines invoke the First Amendment and promise to keep the removal of users or user-generated content to "the absolute minimum." *Community Guidelines*, PARLER (Nov. 2, 2021), https://parler.com/documents/guidelines.pdf [https://perma.cc/4CMY-UFCL]. The policies are set out in a two-page document that focuses mainly on protecting the platform against illegal activity and nuisances like spam or bots. *See id.* Parler's guidelines are silent on hate speech. *See id.* In sharp contrast, Facebook's standards identify and elaborate upon three separate "tiers" of hate speech. *Hate Speech*, META, https://transparency.fb.com/policies/community-standards/hate-speech/ [https://perma.cc/E6WV-2FVH].

[382] *See, e.g.*, Nicole Buckley & Joseph S. Schafer, *'Censorship-Free' Platforms: Evaluating Content Moderation Policies and Practices of Alternative Social Media*, 4 FOR(E)DIALOGUE 2 (2022) (noting that when pressured by Apple's App Store and Google's Play Store, Parler, Bitchute, Gab, and Gettr "were forced to create or adapt . . . content moderation policies"); Jessica Melugin, *Conservative Social Media Platforms Can't Succeed Without Content Moderation*, ORANGE CNTY. REG. (July 22, 2021), https://www.ocregister.com/2021/07/22/conservative-social-media-platforms-cant-succeed-without-content-moderation/ [https://perma.cc/6LUG-KLGN] (noting that lax content moderation can quickly turn conservative platforms into "a hellscape of imposter accounts, offensive memes and pornography").

[383] NetChoice v. Att'y Gen., 34 F.4th 1196, 1229 (11th Cir. 2022) (holding the sequential order requirement "would prevent platforms from expressing messages through post-prioritization and shadow banning").

recommendation algorithms to curate content online is expressive, and thus protected. We disagree.

Recall, content moderation detects and removes (or limits access to) objectionable pieces of content because they violate a firm's policies. Recommendation algorithms order and amplify pieces of content by predicting their relevance to specific users. The former is a content-based activity par excellence that raises serious—if not insurmountable—First Amendment concerns. The latter is not because these algorithms score content according to how likely it is to optimize user engagement, making them content-neutral.

In light of these differences, it is wrong to treat content moderation and content ranking as equally expressive of a platform's outlook. Content moderation is expressive in the sense that it requires the formulation of policies, human oversight of sensitive and nuanced content to ensure that context is properly accounted for, and judgment calls that reflect the "values and viewpoints" of a platform's senior management.[384] But those values and viewpoints have little or no bearing on content-ranking decisions. For instance, for each of the several billion persons on Facebook, ranking algorithms evaluate "thousands of signals . . . to determine what that person might find most relevant."[385] It stands to reason that Facebook's values and viewpoints have little predictive value in determining what a given Facebook user might find relevant. In a nutshell, why would Facebook's senior management's core values matter to an algorithm trying to predict whether a particular user is more engaged by dog photos or cat photos, or by Ezra Klein or Steve Bannon?

Moreover, given the scale and variability of content ranking, it is difficult to intelligibly infer from it any clear expression. As noted, Facebook's ranking algorithms generate relevancy scores in real time for two billion daily users by winnowing thousands of posts into a smaller, personalized list of relevant content.[386] Is it even possible to infer the values and viewpoints of Facebook from this set of newsfeeds for all Facebook users? Obviously not—there are simply too many users (and ordered lists) to make any sense out of all of them. In other words, the probability is very high that one list prioritizes cute cats, another list prioritizes mean dogs, another list shows all things MAGA, another list shows only what Alexandria Ocasio-Cortez is up to, and so on for two billion newsfeeds, in hundreds of languages and an even larger number of cultures and sub-cultures. At this scale, it is simply

---

[384] *See, e.g.*, Will Dunn, *Why Facebook's Future Depends on Nick Clegg*, NEW STATESMAN (Feb. 23, 2022), https://www.newstatesman.com/science-tech/big-tech/2022/02/why-facebooks-future-depends-on-nick-clegg [https://perma.cc/QAH3-DLAU] (describing Meta's new president of global affairs, as "in charge of the political positions and interactions of the social giant").

[385] Lada et al., *supra* note 353.

[386] *See supra* text accompanying note 357.

impossible to derive any coherent set of values or viewpoints on the basis of recommendations.

At best, we can say that these ranking algorithms reflect Facebook's stated goal of delivering relevant newsfeed to all of its users.[387] We might also suggest that they reflect algorithmic predictions optimized by Facebook to determine what would retain users on the platform the longest or otherwise make users engage in activities that would generate more revenue for Facebook. But understanding those actions as exercising editorial judgment, or intently manifesting a specific message or content, is an overreach. It ignores the major differences between humans engaged in editorial judgment and machine-learning algorithms that optimize for engagement by evaluating millions of items based on thousands of factors. Call this an editorial judgment if you wish. But then so is the *New York Times*' slogan "All the News That's Fit to Print." And that slogan stops far short of editorial judgment, which as *Miami Herald* teaches, requires more than just an open-ended commitment to reporting whatever "fits."[388]

3. Benefits of Regulating Algorithmic Ranking

The preceding analysis of the regulation of algorithmic ranking paves a new path for regulating online PHM. It suggests that regulation of a platform's amplification mechanism can survive First Amendment scrutiny, particularly for a compelling government interest such as confronting online PHM.[389] Consider two of the legislative reforms and proposals. According to the sponsor of S. 2024, the harms associated with amplification include "political polarization, social isolation, and addiction" as well as "the algorithmic promotion of abusive, divisive, and extremist content."[390] The bill seeks to combat these harms by providing social media users with greater transparency about algorithmic ranking systems and to offer them the choice of abandoning algorithmically curated experience and associated "filter bubbles" in favor of a chronological newsfeed.[391] Arguably, this would reduce the distribution and impact of misinformation generally (including online

---

[387] Keller argues that "[p]latforms . . . 'speak' through ranking decisions . . . sa[ying] things like 'I predict that you'll like this' or 'I think this is what you're looking for.'" *See* Keller, *Amplification and Its Discontents*, *supra* note 327, at 247. Even if this is speech, it does not amount to an expression of the platforms' values or editorial judgments. *Id.*

[388] Miami Herald Publ'g Co. v. Tornillo, 418 U.S. 241, 258 (1974) (noting that editorial judgment requires "decisions made as to limitations on the size and content of the paper, and treatment of public issues and public officials").

[389] *See supra* Section II.B; Does 1-6 v. Mills, 16 F.4th 20, 32 (1st Cir. 2021), *cert. denied sub nom.* Does 1-3 v. Mills, 142 S. Ct. 17 (2021) ("Stemming the spread of Covid-19 is . . . a compelling interest." (quoting Roman Cath. Diocese of Brooklyn v. Cuomo, 141 S. Ct. 63, 67 (2020)).

[390] Press Release, John Thune, U.S. Senator for S.D., Thune, Colleagues Reintroduce Bipartisan Bill to Increase Internet Platform Transparency (June 10, 2021), https://www.thune.senate.gov/public/index.cfm/2021/6/thune-colleagues-reintroduce-bipartisan-bill-to-increase-internet-platform-transparency [https://perma.cc/R7U7-5FQB].

[391] *Id.*

PHM) by disrupting virality[392] and curation that nudges users towards extreme content,[393] and perhaps mitigating the "illusory truth" effect.[394] On the other hand, when Facebook experimented with this approach by turning off the newsfeed algorithms for some users and substituting a chronological feed, users were not happy, suggesting that few users would exercise their opt-out rights even if given the choice.[395] Obviously, much would depend on the specific implementation of alternative forms of managing and ordering content and whether they provided users with easy-to-use tools that delivered a desirable experience while avoiding unintended consequences.

Another promising avenue for regulation of algorithmic amplification would require friction and middleware. Friction-by-design regulation slows the velocity of viral sharing by imposing various sorts of delays.[396] Hence, friction is a particularly important feature to confront misinformation. "Platform mechanisms that make it easy and frictionless to reshare content will tend to give broader distribution to misinformation."[397] And friction forces platforms to bear the burden of implementing new design requirements, rather than expecting users to understand how curation works or whether they are better off with chronological sorting.

Middleware regulation would require that platforms allow users to replace or modify the platform's built-in ranking algorithms. For example, fact checking organizations or other trusted third-parties might develop competing algorithms that optimize for accuracy and credibility in news stories and penalize reporting that incorporates unfounded rumors and conspiracy theories. Users would decide for themselves which organizations they trust, while platforms would have to modify their architecture to allow those actors to operate. As Fukuyma explains: "At one extreme, middleware could take over the entire user interface of a Facebook or Google, relegating those platforms to the status of 'dumb pipes' that simply serve up raw data, much like the telephone companies. At the other extreme, middleware could operate with a light touch, labeling but otherwise not affecting the content-curation decisions being made by the platforms."[398] Like data portability, which requires platforms to develop common technical standards to enable users to download and export their data to a competitor's service,[399] middleware solu-

---

[392] Algorithmic ranking optimizes for user engagement, *see generally* Chan et al., *supra* note 79; chronological ordering does not.

[393] *See* Vynck et al., *supra* note 98.

[394] *See generally* Section II.B.

[395] *See* Shoshana Wodinsky, *Why the New Big Tech Anti-Algorithm Bill Is Doomed Even if It Succeeds*, Gizmodo (Nov. 10, 2021), https://www.gizmodo.com.au/2021/11/why-the-new-big-tech-anti-algorithm-bill-is-doomed-even-if-it-succeeds/ [https://perma.cc/T4TU-U2CJ].

[396] *See supra* text accompanying note 337.

[397] *Misinformation Amplification Analysis and Tracking Dashboard*, *supra* note 86 (also noting that Facebook having more friction than Twitter to sharing posts explains why Twitter has more misinformation).

[398] *See* Fukuyama, *supra* note 336, at 42.

[399] *See* Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and

tions are also designed to reduce the power of platforms and encourage competition.[400]

Again, it is not our goal to determine which of these alternatives is most likely to succeed. For now, we are satisfied with pointing out that this new path is open, promising, and worthwhile—despite the awesome shadow of the First Amendment. We are confident that content-neutral regulation of ranking algorithms helps combat online misinformation generally including online PHM.

VI. Conclusion

Online public health misinformation is a considerable public health problem. PHM, specifically about COVID-19, has been spreading wildly on platforms, despite their efforts to confront it. Existing legal paths are too narrow or too restricted by the First Amendment to adequately address this problem. Additionally, relying on platforms—private actors—to address this grave social problem has many normative and political shortcomings. Against this background, this Article charted a path forward. It discussed a set of soft-regulation approaches that have been used by other states and in other contexts, and suggested that existing First Amendment doctrine could accommodate stricter regulation of a crucial part of online PHM—content amplification.

The soft-regulation schemes we suggested included adoption of codes of conduct and voluntary enforcement. Both schemes are already being applied in other countries, are well-suited for regulating online speech and misinformation, and have considerable benefits over the existing approach in the United States. And both can be easily and effectively implemented in the United States.

Our argument that some regulation of online misinformation can survive the First Amendment hinges on a distinction between algorithmic amplification (i.e., recommendation algorithms) and content moderation. We explained the technological foundations of this distinction and pointed out some of its legal implications. We think this distinction is important and can serve future research on online speech regulation. The following table sums up this distinction:

on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) Art. 20.

[400] *See supra* text accompanying note 336.

	Recommendation Algorithms	**Content Moderation**
The Procedure	Analyzes thousands of factors to select content that optimizes a pre-defined utility function for each user (typically, user engagement)	Matches content to predefined categories
Main Output	Distributes and amplifies content	Removes or leaves-up content
Content Based or Neutral	Content-neutral. Only marginally relies on the meaning of the content to decide the output	Content-based. Relies heavily on the meaning of the content to decide on the output
Editorial Judgment	Rarely. Amplifies and distributes different content to different users based on the utility function; does not coherently express any viewpoint	Typically. Platforms make decisions about values and procedures that in turn determine the availability of types of content, and thereby differentiating themselves from other platforms
Possible Regulation	Laws that require friction or prohibit shadow-banning	Laws that prohibit sex-trafficking ads, (some) pornography, terrorist speech, etc.
Possible Soft-Regulation	Codes of conduct guiding how to de-amplify hate speech; monitoring and reporting requirements for efforts to add friction to online PHM	Voluntary enforcement to remove terrorist content

Our solutions are not normatively perfect, nor will applying them immediately solve the problem of online PHM. However, this menu of options is much better than any existing legal framework to confront PHM. Our solutions recognize that platforms' policies, and specifically their decisions on which content to amplify, are crucial elements in confronting online PHM. Moreover, and against existing understanding, we also explained that those solutions can be applied now, even under the existing legal doctrines. Using soft-regulation and a close reading of the First Amendment as it pertains to the relevant technology, we charted a path for governments to influence platforms' regulation of PHM.

Importantly, the discussions above feature ways for governments to influence platforms that govern online speech. Thus, these methods could also be used to confront other kinds of online misinformation. In this Article, we argue for applying those methods to confronting PHM. Additional normative, political, and epistemic arguments might be needed to justify using those methods with regard to other kinds of misinformation. But those discussions could surely benefit from the elaborate analysis and justification of using those methods to confront online PHM.

Obviously, this Article leaves many open questions—about the future of soft-regulation and government's levers over platforms, about content amplification as content-neutrality, and about other kinds of harmful online misinformation. These will be discussed, we hope, in future scholarship.