Washington Law Review

Volume 97 | Number 4

12-1-2022

Disorderly Content

Ari Waldman
Northeastern University, a.waldman@northeastern.edu

Follow this and additional works at: https://digitalcommons.law.uw.edu/wlr

Part of the Computer Law Commons, Public Policy Commons, and the Science and Technology Policy Commons

Recommended Citation
Ari Ezra Waldman, Disorderly Content, 97 Wash. L. Rev. 907 (2022).

This Article is brought to you for free and open access by the Washington Law Review at UW Law Digital Commons. It has been accepted for inclusion in Washington Law Review by an authorized editor of UW Law Digital Commons. For more information, please contact lawref@uw.edu.

DISORDERLY CONTENT

Ari Ezra Waldman*

Abstract: Content moderation plays an increasingly important role in the creation and dissemination of expression, thought, and knowledge. And yet, throughout the social media ecosystem, nonnormative and LGBTQ+ sexual expression is disproportionately taken down, restricted, and banned. The current sociolegal literature, which focuses on content moderation as a whole and sees echoes of formal law in the evolution of its values and mechanics, insufficiently captures the ways in which those principles and practices are not only discriminatory, but also resemble structures of power that have long been used to police queer sexual behavior in public spaces.

This Article contributes to the sociolegal literature by approaching content moderation from an explicitly queer perspective, bridging siloed scholarship on law, technology, and LGBTQ+ history. It argues that content moderation for "sexual activity" is an assemblage of social forces that encodes queerness as sexual in a way that straightness is not. This is the case because far from simply reflecting free speech principles, as several scholars have argued, content moderation in fact resembles oppressive anti-vice campaigns from the middle of the last century in which "disorderly conduct," "vagrancy," "lewdness," and other vague morality statutes were disproportionately enforced against queer behavior in public. This analogy highlights underappreciated pieces of the content moderation puzzle. Like anti-vice campaigns, sexual content moderation emerged from similar sociolegal contexts, relies on similar justificatory discourses, leverages similarly vague rules, similarly operates mostly without rigorous, science-backed expertise in sexual content, also disproportionately silences queer content, and similarly does so without due process. Ultimately, I argue that like anti-vice enforcement, sexual content moderation results in the maintenance and reification of social media as "straight spaces" that are hostile to queer, nonnormative expression.

This Article provides a full, critical account of sexual content moderation and its effects on queer expression. It details and challenges the current content moderation literature and explores potential new directions for scholarship, moderation, and law. The similarities and differences between anti-vice enforcement and sexual content moderation also suggest a way forward, offering novel justifications for modest legal reform, social activism, and platform responsibility.

---

* Professor of Law & Computer Science and Faculty Director, Center for Law, Information, and Creativity, Northeastern University. PhD, Columbia University; JD, Harvard Law School; AB, Harvard College. Special thanks to Hadar Aviram, Jack Balkin, Hannah Bloch-Wehba, Ryan Calo, Julie Cohen, Rashmi Dyal-Chand, Veena Dubal, Richard Deynard, Niklas Eder, Bill Eskridge, Tarleton Gillespie, Samantha Godwin, Nikolas Guggenberger, Woodrow Hartzog, Bonnie Kaplan, Asaf Lubin, Anna Lvovsky, Dwight McBride, Doug NeJamie, Helen Nissenbaum, Jennifer Oliva, Jeremy Paul, Sonia Rolland, Blaine Saito, Eden Sarid, Reuel Schiller, Jodi Short, Scott Skinner-Thompson, Alicia Sollow-Niederman, Lucy Williams, Patricia Williams, and the participants in the Yale Information Society Project and the NYU Privacy Research Group. Maya Bornstein and Meg Foster provided essential research assistance. All disorderly content—syntactical, grammatical, and substantive errors—is my own.

INTRODUCTION 909
I. THE CONTENT MODERATION ASSEMBLAGE 915
A. Deconstructing Moderation 916
B. Putting Moderation Back Together 921
II. A HISTORICAL MODEL FOR SEXUAL CONTENT MODERATION 923
A. Legal Contexts, Discourses, and Values 924
1. Repressive Origins 925
2. Discourses Rationalizing Repression 928
a. Economic Rationales 928
b. Reasserting Institutional Control 929
c. Protecting the Children and Other Vulnerable Populations 931
B. The Assemblages' Mechanics 934
1. Vague Rules 934
2. Technological and Human Expertise 937
3. Disproportionate Antiqueer Enforcement 940
a. Asymmetrical Enforcement: Data and Experiences 940
b. Community Enforcement 942
c. Intersectional Injustice 944
4. Without Due Process 946
C. Sexual Content Moderation's Path Dependencies 948
D. Limitations 950
III. SOCIAL MEDIA AS A "STRAIGHT SPACE" 951
A. Stigmatizing Queerness 953
B. Technology in Straight Spaces 957
C. The Role of Law 959
IV. CREATING QUEER INCLUSIVE SEXUAL CONTENT MODERATION 963
A. The Insufficiency of Current Proposals 964
B. Lessons from Anti-Vice Policing 968
1. Evidence, Not Symbols 969
2. Public Governance, Not False Transparency 971
3. Anti-Subordination, Not Blanket Immunity 974
CONCLUSION 976

INTRODUCTION

Twitter identifies drag queen accounts as more toxic that those of white supremacists.[1] YouTube hid a video entitled "I Am Transgender," restricted another where a content creator talks about her bisexuality, and demonetized a slew of LGBTQ videos as controversial.[2] Instagram takes down posts by the queer-focused Leslie-Lohman Museum, restricts the reach of queer sex workers and porn stars, and has removed pictures of shirtless trans men that would be perfectly anodyne if the subject were cisgender.[3] The very process of flagging—through which other members report allegedly offending content to platforms—censors queer content by subjecting it to the heteronormative judgments of others.[4]

In many cases of queer censorship online that gain enough attention in the press, platforms eventually apologize and concede that their algorithm made mistakes.[5] Legal scholarship on content moderation—the sociotechnical rules and systems that decide the fate of user-generated content on digital platforms—also tends to focus on mistakes: the content

---

1. Thiago Dias Oliva, Dennys Marcelo Antonialli & Alessandra Gomes, *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, 25 SEXUALITY & CULTURE 700, 703 (2021).

2. Libby Watson, *YouTube's Restricted Mode Is Hiding Some LGBT Content*, GIZMODO (Mar. 18, 2017), https://www.gizmodo.com.au/2017/03/youtubes-restricted-mode-is-hiding-some-lgbt-content [https://perma.cc/WR4Z-RAXT]; Molly Priddy, *Why Is YouTube Demonetizing LGBTQ Videos?*, AUTOSTRADDLE (Sept. 22, 2017), https://www.autostraddle.com/why-is-youtube-demonetizing-lgbtqia-videos-395058/ [https://perma.cc/Z8Y3-NKBN].

3. Leslie-Lohman Museum of Art (@leslielohmanmuseum), INSTAGRAM (June 10, 2021), https://www.instagram.com/p/CP83f34F3LS/?utm_medium=copy_link (last visited Oct. 13, 2022); Eli Erlick, *How Instagram May Be Unwittingly Censoring the Queer Community*, THEM (Jan. 30, 2018), https://www.them.us/story/instagram-may-be-unwittingly-censoring-the-queer-community [https://perma.cc/Y7W3-A9E4]; Cherie DeVille, *Think Conservatives Are Censored on Social Media? Try Being a Porn Star*, DAILY BEAST (May 30, 2021), https://www.thedailybeast.com/think-conservatives-are-censored-on-social-media-try-being-a-porn-star [https://perma.cc/RG9L-Z3NS]; Richard Renaldi (@renaldiphotos), INSTAGRAM (July 1, 2021), https://www.instagram.com/p/CQynMmIBCk9/?utm_medium =share_sheet (last visited Oct. 13, 2022).

4. Kate Crawford & Tarleton Gillespie, *What Is a Flag For? Social Media Reporting Tools and the Vocabulary of Complaint*, 18 NEW MEDIA & SOC'Y 410, 411, 413 (2016).

5. *E.g.*, Susan Wojcicki, *A Message on Pride and LGBTQ Initiatives*, YOUTUBE OFF. BLOG (June 19, 2017), https://youtube-creators.googleblog.com/2017/06/a-message-on-pride-and-lgbtq-initiatives.html [https://perma.cc/52ED-EWMH]; Lily Wakefield, *TikTok Insists It 'Celebrates and Protects' Queer Creators but Apologises for Censoring 'Vulnerable' LGBT+ Users*, PINKNEWS (Sept. 23, 2020), https://www.pinknews.co.uk/2020/09/23/tiktok-censorship-lgbt-content-theo-bertram-parliamentary-sub-committee/ [https://perma.cc/4XV8-HX94]; EJ Dickson, *Why Did Instagram Confuse These Ads Featuring LGBTQ People for Escort Ads?*, ROLLING STONE (July 11, 2019), https://www.rollingstone.com/culture/culture-features/instagram-transgender-sex-workers-857667/ [https://perma.cc/NW4R-836V]; *see also* Johanna Wright, *An Update on Restricted Mode*, YOUTUBE OFF. BLOG (Apr. 21, 2017), https://blog.youtube/news-and-events/an-update-on-restricted-mode/ [https://perma.cc/XL5V-ASKN].

that should never have been taken down or should never have been allowed in the first place.[6] Other scholars have studied content moderation's animating values, as well as its inequities, inadequate sociotechnical processes, and harmful effects on those doing the monitoring.[7] In other words, it is common to hear about content moderation going wrong. A troubling, but more compelling reading of this pattern of restrictions is that censoring queer content is not content moderation going wrong at all, but rather the natural consequence of sexual content moderation doing exactly what it was designed to do.

The disproportionate removal of some content and not others must be studied for what it reveals about platform power, free expression, and regulation. This Article contributes to our understanding of content moderation by approaching it from the experiences of queer people and those whose sexual expression sits outside traditional norms. Simply put, I argue that sexual content moderation is best understood as an assemblage of social, legal, and technological forces that encodes queerness as inherently sexual in a way that heteronormative content is not. In that way, sexual content moderation does not reflect free speech principles, as many scholars have argued; rather, it resembles anti-vice

6. *See, e.g.*, Evelyn Douek, *Governing Online Speech: From "Post-as-Trumps" to Proportionality and Probability*, 121 COLUM. L. REV. 759, 762, 808 (2021) [hereinafter Douek, *Governing*] ("There is no denying, and indeed Facebook acknowledged, that these were mistakes.").

7. On content moderation's animating values, see, for example, TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA (2018); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018) [hereinafter Klonick, *Governors*]; Douek, *Governing*, *supra* note 6. On content moderation's inequities, see, for example, GILLESPIE, *supra*; Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870 (2019) [hereinafter Citron, *Sexual Privacy*]; Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353 (2017); Ariadna Matamoros-Fernández, *Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube*, 20 INFO. COMMC'N & SOC'Y 930 (2017); Sarah T. Roberts, *Digital Detritus: 'Error' and the Logic of Opacity in Social Media Content Moderation*, 23 FIRST MONDAY (2018) [https://perma.cc/J8J7-KFRT] [hereinafter Roberts, *Detritus*]; Julia Angwin & Hannes Grassegger, *Facebook's Secret Censorship Rules Protect White Men from Hate Speech But Not Black Children*, MOTHER JONES (June 28, 2017), https://www.motherjones.com/politics/2017/06/facebooks-secret-censorship-rules-protect-white-men-from-hate-speech-but-not-black-children/ [https://perma.cc/TC94-EXHX]. On content moderation's technical elements, see, for example, SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA (2019) [hereinafter ROBERTS, SCREEN]; Hannah Bloch-Wehba, *Automation in Moderation*, 53 CORNELL INT'L L.J. 41 (2020); Crawford & Gillespie, *supra* note 4; J. NATHAN MATIAS, AMY JOHNSON, WHITNEY ERIN BOESEL, BRIAN KEEGAN, JACLYN FRIEDMAN & CHARLIE DETAR, REPORTING, REVIEWING, AND RESPONDING TO HARASSMENT ON TWITTER (2015), https://arxiv.org/pdf/1505.03359.pdf [https://perma.cc/2HEY-JQG6]. On harms to moderators, see, for example, MARY L. GRAY & SIDDHARTH SURI, GHOST WORK (2019); Sarah T. Roberts, *Commercial Content Moderators: Digital Laborer's Dirty Work*, *in* THE INTERSECTIONAL INTERNET: RACE, SEX, CLASS AND CULTURE ONLINE 147–60 (Safiya Noble & Brendesha Tynes eds., 2016) [hereinafter Roberts, *Dirty Work*].

policing and, as a result, ultimately reifies social media as "straight spaces." The implications of this analogy are profound: Sexual content moderation was not designed with express anti-queer intentions, but just like with anti-vice policing, the disproportionate regulation of queer sexuality is the direct result of the institution's constituent pieces working as they were designed.

There are three elements to this argument. First, content moderation is an assemblage—a convergence of social phenomena that exerts power in ways that each individual element could not—comprising background law, values, rules, design, technology, and people that come together to achieve platforms' economic, moral, and political goals.[8] Scholars have studied many of these pieces independently. Part I reconstructs that literature to demonstrate the different yet complementary roles played by each piece.

The second part of the argument is that sexual content moderation resembles anti-vice policing of the mid-twentieth century. This argument is a historical one; that is, we have seen these precise pieces assembled before with similar consequences for public expression of queer sexuality. Part II describes how sexual content moderation parallels the assemblage of sociolegal contexts, discourses, applicable rules, technologies and expertise, enforcement strategies, and processes that policed queer behavior from the 1930s through the 1960s, where police used anti-vice laws like "disorderly conduct," "lewdness," and "vagrancy" to harass and arrest those exhibiting nonnormative sexual behavior and to shut down those bars that served queer patrons.[9] Although we should be careful not to make too much of this analogy,[10] the parallels between anti-vice policing and content moderation are unmistakable. History may not always repeat, but it certainly rhymes.[11]

---

8. GILLES DELEUZE & FÈLIX GUATTARI, A THOUSAND PLATEAUS (1987); MANUEL DELANDA, A NEW PHILOSOPHY OF SOCIETY: ASSEMBLAGE THEORY AND SOCIAL COMPLEXITY (2006).

9. ANNA LVOVSKY, VICE PATROL 3–8 (2021). These elements are not always distinct. Several progressive legal movements recognize that sociological, institutional, and discursive contexts cannot be disaggregated from law, itself an endogenous creation of interested social and economic actors. *E.g.*, MORTON J. HORWITZ, THE TRANSFORMATION OF AMERICAN LAW, 1870-1960: THE CRISIS OF LEGAL ORTHODOXY (1994); Mark Tushnet, *Critical Legal Studies: A Political History*, 100 YALE L.J. 1515, 1516–17 (1991); Jedediah Britton-Purdy, David Singh Grewal, Amy Kapczynski & K. Sabeel Rahman, *Building a Law-and-Political-Economy Framework: Beyond the Twentieth-Century Synthesis*, 129 YALE L.J. 1784, 1792–94 (2020). The assemblage construct asks us to see how even overlapping elements work together exert power in unique ways. Kevin D. Haggerty & Richard V. Ericson, *The Surveillant Assemblage*, 51 BRIT. J. SOCIO. 605, 608 (2000).

10. *See infra* section II.C and discussions of differences between anti-vice policing and sexual content moderation throughout sections II.A and II.B.

11. A version of this aphorism is commonly attributed to Mark Twain, but there is no evidence he

This analogy teaches us something that has been missing in the content moderation literature. It is not enough to understand content moderation as a pragmatic balancing of competing interests.[12] Nor is it sufficient to conceptualize it as ex ante administration rather than adjudication after the fact.[13] We should also see sexual content moderation as yet another battlefield in a long struggle over the visibility of queer sexuality. This is the third piece of my argument. Through its similarities and differences, the analogy to anti-vice policing surfaces three ways that sexual content moderation encodes queerness as illicit sexuality and makes and maintains social media platforms as "straight spaces," or spaces built on heterosexual norms, constructed to be unwelcome to queer and nonnormative expression, and designed to reify the heteronormative supremacy of our institutions.[14]

First, the effect of sexual content moderation's on-the-ground practices, like the practical effect of anti-vice policing, is to deny queer people dignity and to chill nonnormative sexual expression. This skews the production and dissemination of public knowledge about queer life. Second, sexual content moderation and anti-vice policing use the technologies of their day to disproportionately surveil nonnormative sexual behavior, imbuing technology with heteronormative politics. Finally, the analogy also demonstrates how law contributes to the maintenance of straight spaces online. In both contexts, the law amplified the power of those regulating queer behavior.[15] However, unlike victims of anti-vice laws, who were sometimes able to use the courts as weapons against police overreach,[16] queer people online have no access to the kind of judicial discretion and impact litigation that ultimately protected victims of anti-vice policing. That asymmetry threatens to maintain structures of power in ways anti-vice authorities could never have fathomed.

---

said it. He did say "[h]istory never repeats itself" in 1874. SAMUEL CLEMENS & CHARLES WARNER, THE GILDED AGE: A TALE OF TO-DAY 430 (1874). The earliest known recorded version of the adage comes from the psychoanalyst Theodor Reik. THEODOR REIK, CURIOSITIES OF THE SELF: ILLUSIONS WE HAVE ABOUT OURSELVES 133 (1965).

12. Douek, *Governing*, *supra* note 6, at 763.

13. Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. (forthcoming 2022) (manuscript at 4) (on file at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4005326 [https://perma.cc/7MEC-LCEW]) [hereinafter Douek, *Systems*].

14. The notion of "straight space" is adapted from the notion of "white space" in critical race theory. Elijah Anderson, *"The White Space"*, 1 SOCIO. RACE & ETHNICITY 10, 10 (2015).

15. LVOVSKY, *supra* note 9, at 5–6; JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 75–107 (2019) (demonstrating how information industry actors leverage legal tools to amplify their power and insulate themselves from liability).

16. WILLIAM N. ESKRIDGE, JR., GAYLAW: CHALLENGING THE APARTHEID OF THE CLOSET 101–11 (2002) [hereinafter ESKRIDGE, GAYLAW]; LVOVSKY, *supra* note 9, at 134–36, 257–58.

Now is the time to deepen our understanding of content moderation. Platforms are under unprecedented scrutiny for anti-trust, competition, and privacy law violations.[17] Content moderation itself has caught the eye of policymakers across the political spectrum.[18] Society's overdue attention to racial, gender, and sexual justice is taking a magnifying glass to platforms' complicity in reifying traditional structures of power.[19] And, at the same time, some of the biggest social media platforms today are building social spaces for the future and are, therefore, at risk of making the same mistakes of the past.[20] In other words, cascading social forces threaten to change the face of social media and platform governance. Making social media queer friendly after this period of contestation requires a more nuanced, intersectional understanding of the institutions at play and, ultimately, sustained social activism in the face of a rigid sexual hierarchy.

This Article proceeds in four Parts. Part I reviews the literature on content governance, making the descriptive claim that content moderation is best understood as an assemblage of ideas, rules, technology, and people rather than as a monolithic institution. This section concludes by highlighting three gaps in the literature—namely, the lack of granularity in discussions of content moderation, the lack of explicitly queer examples, and the near-universal focus on moderation mistakes as case studies for broader conclusions about content moderation as a whole. Part II argues that the assemblage of sexual content moderation is similar to the assemblage of anti-vice policing in the United States from the 1930s through the 1960s, demonstrating that the values, machinery, and rationales of the former live on in the latter. Part III demonstrates how the analogy to anti-vice policing highlights sexual content moderation's role in sustaining social media as straight spaces hostile to queer people and nonnormative sexual discourse. Part IV considers proposals to regulate platform moderation through a queer lens and then looks to the anti-vice

17. *E.g.*, Cecilia Kang & David McCabe, *House Lawmakers Are Considering 6 Bills Aimed at Big Tech*, N.Y. TIMES (June 23, 2021), https://www.nytimes.com/2021/06/23/technology/big-tech-antitrust-bills.html [https://perma.cc/M2G9-PB38].

18. David Morar & Bruna Martins dos Santos, *The Push for Content Moderation Legislation Around the World*, BROOKINGS (Sept. 21, 2020), https://www.brookings.edu/blog/techtank/2020/09/21/the-push-for-content-moderation-legislation-around-the-world/ [https://perma.cc/L4PD-NTA8].

19. *E.g.*, Charlton McIlwain, *Of Course Technology Perpetuates Racism. It Was Designed That Way.*, MIT TECH. REV. (June 3, 2020), https://www.technologyreview.com/2020/06/03/1002589/technology-perpetuates-racism-by-design-simulmatics-charlton-mcilwain/ [https://perma.cc/RY8H-GVDG].

20. *E.g.*, James Factora, *The Metaverse Is Going to Suck for Queer People*, THEM (Feb. 9, 2022), https://www.them.us/story/metaverse-queer-lgbtq-users-harassment-discrimination-facebook [https://perma.cc/WCT7-27JN].

policing context to suggest several steps that could make social media more inclusive of queer expression. The Article briefly concludes by situating its contributions in a larger research agenda about the conflict among technology, law, and queer liberation.

Before turning to my analysis, let me define several terms. "Queer" is a flexible term that refers to disrupting narratives and institutions structured around traditional understandings of sex and gender.[21] To be queer is to be nonnormative.[22] This Article uses "queer" to capture people and behaviors that do not conform to traditional sexual norms. Under this definition, queer can capture content related to sex work, pornography, sex toys, nonmonogamy, and other nonheteronormative uses of our bodies. Although I recognize that queer is not synonymous with the lesbian, gay, bisexual, and transgender community, I will also use "queer" as an umbrella term for these sexual and gender minorities for ease and fluidity of language.

Because nonnormative sexuality is one defining feature of queerness, this Article focuses its analysis on *sexual content moderation* rather than on content moderation as a whole. Sexual content moderation is a subset of moderation rules and procedures that evaluate content of a sexual nature, including sexual activity, nudity, pornography, intimacy, and the like. All major platforms discussed below have specific policies for "sexual activity."[23] I will also refer to those major platforms as *social media*—namely, those internet platforms that allow and derive value from the creation and exchange of user-generated content.[24] Social media platforms are a subset of the Internet as a whole; the largest of those

---

21. *See, e.g.*, EVE KOSOFSKY SEDGWICK, TENDENCIES 8 (1993) (defining "queer" as "the open mesh of possibilities . . . of meaning when the constituent elements of anyone's gender, of anyone's sexuality, aren't made (or *can't be* made) to signify monolithically") (emphasis in original); *see also* Lauren Berlant & Michael Warner, *What Does Queer Theory Teach Us About X?*, 110 PMLA 343, 346 (1995); JUDITH BUTLER, BODIES THAT MATTER 177 (Routledge Classics ed. 2011) (defining "queer" as "a contestation of the terms of sexual legitimacy").

22. DAVID M. HALPERIN, SAINT FOUCAULT: TOWARDS A GAY HAGIOGRAPHY 62 (1995) (defining "queer" as "whatever is at odds with the normal, the legitimate, the dominant. . . . It . . . demarcates not a positivity but a positionality vis-à-vis the normative") (emphasis omitted).

23. *E.g.*, *Adult Nudity and Sexual Activity*, META, https://transparency.fb.com/policies/community-standards/adult-nudity-sexual-activity/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2Fadult_nudity_sexual_activity [https://perma.cc/9D42-JE36]; *Adult Nudity and Sexual Activities*, TIKTOK, https://www.tiktok.com/community-guidelines?lang=en#30 [https://perma.cc/5NJG-VTGH]; *Nudity & Sexual Content Policy*, YOUTUBE, https://support.google.com/youtube/answer/2802002#zippy=%2Cother-types-of-content-that-violate-this-policy [https://perma.cc/P9FN-CXSY].

24. Andreas M. Kaplan & Michael Haenlein, *Users of the World, Unite! The Challenges and Opportunities of Social Media*, 53 BUS. HORIZONS 59, 60–61 (2010); Caleb T. Carr & Rebecca A. Hayes, *Social Media: Defining, Developing, and Divining*, 23 ATL. J. COMMC'N 46, 47–48 (2015).

platforms are a further subset. Most of this Article's examples come from today's major platforms, including Facebook, Instagram, YouTube, Twitter, and TikTok.

I. THE CONTENT MODERATION ASSEMBLAGE

An assemblage is a convergence of discrete systems of control.[25] The philosophers Gilles Deleuze and Felix Guattari stated that assemblages are a "multiplicity of heterogenous objects, whose unity comes solely from the fact these items function together, that they 'work' together as a functional entity."[26] Assemblages can include people, institutions, things, knowledge, and technologies, just to name a few; they seek to explain "all the voices present within a single voice."[27] Their goal is to exert control, to "introduc[e] breaks and divisions into otherwise free-flowing phenomena,"[28] like, in this case, information.

For instance, city traffic, which could be a dangerous mess if unregulated, is controlled through an assemblage of policymakers, experts, roads, directional signs, traffic lights, norms of cooperation (at a four-way intersection) or rule-breaking (as in the case of jaywalking in New York City), traffic cops, yellow and white solid or dotted lines, speed limits, sensors, and even the availability of alternative forms of transportation like subways. Kevin Haggerty and Richard Ericson used the notion of an assemblage to explain how overlapping forms of surveillance come together to extract data and exert power over individuals in ways disparate, non-linked surveillance tools cannot.[29] Tania Bucher conceptualized Facebook as an assemblage of "the social and technical," including people, nonhumans, algorithms, social practices, and cultural values.[30] In all cases, individual elements have "distinctive but mutually shaping roles."[31]

Content moderation involves the background laws, ideas, rules,

---

25. Haggerty & Ericson, *supra* note 9, at 606.

26. Paul Patton, *Metamoropho-Logic: Bodies and Powers in* A Thousand Plateaus, 25 J. BRIT. SOC'Y PHENOMENOLOGY, 157, 158 (1994).

27. Deleuze & Guattari, *supra* note 8, at 88. They also need to explain the effects of missing voices. *See, e.g.*, Judy Wajcman, *Reflections on Gender and Technology Studies: In What State Is the Art?*, 30 SOC. STUD. SCI. 447, 452 (2000) (highlighting the blind spot of social research that follows the work of social actors without considering the work of those excluded from sociopolitical and scientific spaces for structural reasons).

28. Haggerty & Ericson, *supra* note 9, at 608.

29. *Id.*

30. TANIA BUCHER, IF . . . THEN 50–52 (2018).

31. José van Dijck, *Facebook and the Engineering of Connectivity*, 19 INT'L J. RSCH. NEW MEDIA TECHS. 141, 146 (2012).

technologies, and people that collectively—both ex ante and ex post—determine the appropriateness of content for a given platform.[32] Understanding content moderation not as a monolithic institution, but rather as several assemblages of social forces, allows us to see previously hidden patterns. One of those patterns is sexual content moderation's resemblance to anti-vice policing and its designed-in and natural hostility to queer and nonnormative content. This Part recharacterizes relevant sociolegal scholarship on content moderation as an assemblage and concludes by surfacing three prominent gaps in the literature. In particular, scholars have insufficiently interrogated different types of content moderation, the experiences of queer content creators, and the implications of the discourse of mistake.

A. Deconstructing Moderation

There are at least six overlapping pieces to the content moderation puzzle: background law, discourses, rules, design, machines, and people.[33] Content moderation—and our understanding of its strengths and limitations—would be incomplete without any one of them. And although these elements change, overlap, and influence each other, they each "assemble" to create a system in which platforms control the flow of content in ways that no single force could.[34] Background law enables corporate control over their platforms. Law's expressive value amplifies the politics and discourses of content moderation. Rules known as "community guidelines" implement those discourses into practice. Technology design frames and sometimes predetermines user adherence to those rules. Machines implement the rules' requirements on user-generated content. And different groups of people play different roles in moderation, from the workers who design platforms and algorithmic moderation systems to users who flag content they find inappropriate and appeal adverse algorithmic moderation decisions to human moderators.

Background law. Scholars writing about the law and political economy of informational capitalism have shown how platforms leveraged the law

32. *See* Niva Elkin-Koren, Giovanni De Gregorio & Maayan Perel, *Social Media as Contractual Networks: A Bottom Up Check on Content Moderation*, 107 IOWA L. REV. 987, 995–96 (2022); *see also* Douek, *Systems*, *supra* note 13, at *3 & n.2.

33. This formulation evokes early information law scholarship describing internet governance as a product of law, code, markets, and norms. Joel R. Reidenberg, *Lex Informatica*, 76 TEX. L. REV. 553, 554–55 (1998); LAWRENCE LESSIG, CODE VERSION 2.0 (2006). This particular list of constitutive elements builds on the work of Ysabel Gerrard and Helen Thornham. Ysabel Gerrard & Helen Thornham, *Content Moderation: Social Media's Sexist Assemblages*, 22 NEW MEDIA & SOC'Y 1266, 1269 (2020).

34. Haggerty & Ericson, *supra* note 9, at 608.

to place their business models out of reach of public governance.[35] For example, the First Amendment allows platforms to restrict content that violates their standards.[36] Section 230 of the Communications Decency Act immunizes them from lawsuit when they exercise that discretion.[37] Anupam Chander memorably argues that the law made Silicon Valley, insulating the information industry from accountability and allowing companies to collect information, monetize it, and grow without limit.[38] In other words, technology companies used the law to create an information economy with few limits on platform power, including the power to control content. Therefore, content moderation emerged in a legal context in which platform interests would determine the flow of expression without much interference from regulators.

Discourses. Other scholars have explored the discourses, values, and principles that animate content moderation. Early content moderation was heavily influenced by the First Amendment.[39] Scholars argued that the "marketplace of ideas" metaphor—the notion that good speech, like good products, naturally succeed in unregulated spaces governed by supply and demand[40]—dominated platform content governance rules because liberal free speech values like content and viewpoint neutrality were familiar to those lawyers who wrote the rules in the first place.[41] Such discourses also fit well with Silicon Valley's libertarian ethos because of the latter's hostility to hierarchical, top-down government regulation.[42] However, as Evelyn Douek argues, platforms' reliance on liberal free speech norms was untenable. All rules involve making normative choices and platforms' particular choices made them complicit in systemic social harm: genocide, spread of misinformation, and amplification of hate, among others. As a result, Douek argues that platforms replaced speech's primacy with

---

35. COHEN, *supra* note 15, at 7.

36. Eric Goldman, *Of Course the First Amendment Protects Google and Facebook (and It's Not a Close Question)*, KNIGHT FIRST AMEND. INST. (Feb. 26, 2018), https://knightcolumbia.org/content/course-first-amendment-protects-google-and-facebook-and-its-not-close-question [https://perma.cc/7UG4-JXNL].

37. Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 403 (2017).

38. Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639 (2014).

39. GILLESPIE, *supra* note 7, at 40; Klonick, *Governors*, *supra* note 7, at 1621.

40. *See* Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting); Ari Ezra Waldman, *The Marketplace of Fake News*, 20 J. CONST. L. 845, 851–56 (2018); Joseph Blocher, *Institutions in the Marketplace of Ideas*, 57 DUKE L.J. 821, 829–32 (2008).

41. Klonick, *Governors*, *supra* note 7, at 1618–25.

42. *See* EMILY CHANG, BROTOPIA: BREAKING UP THE BOYS' CLUB OF SILICON VALLEY 60–63 (2018); Sheelah Kolhatkar, *The Tech Industry's Gender-Discrimination Problem*, NEW YORKER (Nov. 13, 2017), https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem [https://perma.cc/Z3YB-FVGW].

"systemic balancing" of multiple interests.[43] Under this regime, user speech rights could be limited for legitimate reasons, provided the limits were proportional to the harm.[44] Jonathan Zittrain described this shift in rules from those based on faith in the generativity of platforms to rules reflecting the real-world costs of limitless online speech.[45]

Rules. Gillespie has analyzed those rules extensively. Described in terms of service (TOSs) and "community guidelines," content rules implement platforms' evolving values. TOSs are written like contracts and detail the terms under which individuals and platforms interact.[46] Community guidelines, on the other hand, are written in "deliberately plainspoken language" and describe what kind of content the platform thinks is or is not appropriate for upload.[47] Community guidelines reflect the "character" of the platform and, by implication, of the people who created and use it.[48] They reflect biases, politics, and what platform rule makers think most people—and, importantly, most advertisers—would want to see.[49] And it is not just platforms and mobile apps that have these rules; Apple and Google have community guidelines that apply to apps hosted on their stores.[50]

Design. Independent research agendas focus on how technology implements those community guidelines, particularly through background design and algorithmic moderation. Though not explicitly writing about content moderation, scholars have demonstrated that the content we see online depends first and foremost on the code-based infrastructure of a platform.[51] Whereas rules set the stage for construction, design predetermines what individuals can and cannot do in a space.[52] By way of

---

43. Douek, *Governing*, *supra* note 6, at 763.

44. *Id.* at 784.

45. Jonathan Zittrain, Three Eras of Digital Governance (Sept. 15, 2019) (unpublished manuscript), https://ssrn.com/abstract=3458435 [https://perma.cc/9CNZ-YDS3].

46. GILLESPIE, *supra* note 7, at 46.

47. *Id.*

48. *Id.* at 48.

49. *Id.* at 19, 35.

50. Luis E. Hestres, *App Neutrality: Apple's App Store and Freedom of Expression Online*, 7 INT'L J. COMMC'N 1265, 1266 (2013).

51. *See* BATYA FRIEDMAN & DAVID G. HENDRY, VALUE SENSITIVE DESIGN: SHAPING TECHNOLOGY WITH MORAL IMAGINATION (2019); WOODROW HARTZOG, PRIVACY'S BLUEPRINT 11–14 (2018); MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE (2018); Mary Flanagan, Daniel C. Howe & Helen Nissenbaum, *Embodying Values in Technology: Theory and Practice*, *in* INFORMATION TECHNOLOGY AND MORAL PHILOSOPHY 322 (Jeroen van den Hoven & John Weckert eds., 2008).

52. *See, e.g.*, KIM DOVEY, FRAMING PLACES: MEDIATING POWER IN BUILT FORM 1 (2d ed. 2008) (explaining how built environments mediate, construct, and reinforce power structures); HENRI

analogy, if community guidelines are the municipal construction approvals, zoning rules, licensure regimes, and workplace safety requirements that govern home construction, design is the house's foundation, architecture, engineering, and building blocks. Both are necessary to build the house. In this way, the sociolegal literature arguing that "code is law" and exploring privacy by design is also about content moderation.[53] Design is a prior restraint on content—the first stage of code-based content moderation—because it requires users to follow technical specifications when uploading content: Instagram videos can only be so long and images can only be certain sizes; Twitter restricts individuals to a set number of characters; TikTok is video-only. As Joel Reidenberg presciently noted, design has "rule-making power" because platform architecture and technical capabilities place limits on user content.[54]

Technology. Machines play another role in content moderation. Today, artificial intelligence automatically identifies, classifies, categorizes, and blocks or takes down content before any human sees it, implementing community guidelines through code.[55] Its primacy in content moderation stems from need and know-how: Today's social media platforms host far more content than they used to,[56] and only AI can screen "unprecedented" numbers of images and videos for women's nipples or copyrighted content.[57] As Facebook CEO Mark Zuckerberg noted in 2018, using AI to proactively screen content only became "possible recently because of advances in artificial intelligence."[58] In this case, technology enabled Facebook's scale.

People. That said, algorithms do not moderate alone. In addition to platform designers and the people who write content moderation rules in the first place, scholars have identified three other sets of people that are part of the content moderation assemblage: commercial content moderators, users, and independent overseers. Commercial content moderators clean up after technology: They restore content that should

---

LEFEBVRE, THE PRODUCTION OF SPACE 224 (Donald Nicholson-Smith trans., 1991) (describing "monumental space[s]" as organized spaces in which certain acts may take place and thus, conversely, what acts may *not* take place).

53. LESSIG, *supra* note 33.

54. Reidenberg, *supra* note 33, at 555, 570–71.

55. 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 (2020).

56. Douek, *Governing*, *supra* note 6, at 791.

57. *Id.* at 793–94.

58. Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK, https://www.facebook.com/notes/751449002072082/ (last visited July 2, 2021).

not have been taken down, reconsider automatic moderation decisions, and act on content AI erroneously let slip through the cracks.[59] Ideally, people bring context and executive function to content moderation questions.[60] But they have to do so repeatedly and in mere seconds while following a booklet of general instructions in front of them.[61] As Sarah Roberts explains, many commercial content moderators today are hired from the Global South, as independent contractors by third-party vendors.[62] Their pay is low and their work exposes them to abusive, violent, and gory content, contributing to social and psychological harm.[63] A small additional layer of human moderators employed directly by Facebook can check their work.[64] These layers of people remain essential: there are just certain things machines cannot do.[65]

Platforms' users also engage in content moderation when they flag content they find inappropriate or appeal an algorithmic moderation decision.[66] Flagging is a feature of platform design that allows users to classify content as offensive, inappropriate, or in violation of their perception of the rules.[67] It serves two ends: conscripting the community of users into the process of reviewing content at scale and legitimizing content moderation by making it, at least superficially, a reflection of user preferences.[68] Those users whose content was blocked or restricted ex ante can also click on a button to appeal a machine's decision to a human moderator.[69] Flagged and appealed content is put in a queue, where

59. Adrien Chen, *Inside Facebook's Outsourced Anti-Porn and Gore Brigade, Where 'Camel Toes' Are More Offensive Than 'Crushed Heads'*, GAWKER (Feb. 16, 2012, 3:45 PM), https://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads [https://perma.cc/E22A-TNAF] [hereinafter Chen, *Inside*].

60. Klonick, *Governors*, *supra* note 7, at 1640.

61. Sarah T. Roberts, *Social Media's Silent Filter*, ATLANTIC (Mar. 8, 2017), https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/ [https://perma.cc/3A8C-ZQGN] [hereinafter Roberts, *Filter*].

62. Roberts, *Detritus*, *supra* note 7; Roberts, *Filter*, *supra* note 61.

63. *See* Roberts, *Dirty Work*, *supra* note 7; Olivia Solon, *Underpaid and Overburdened: The Life of a Facebook Moderator*, GUARDIAN, https://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content [https://perma.cc/2NGG-LRN8]; Complaint for Damages, Soto v. Microsoft Corp., No. 16-2-31049-4 SEA (Wash. Super. Ct. Dec. 30, 2016).

64. Klonick, *Governors*, *supra* note 7, at 1640–41.

65. Frank Pasquale, *Professional Judgment in an Era of Artificial Intelligence and Machine Learning*, 46 BOUNDARY 2, 73 (2019) [hereinafter Pasquale, *Professional*].

66. Crawford & Gillespie, *supra* note 4, at 410.

67. *Id.* at 411.

68. *Id.* at 412.

69. *I Don't Think Facebook Should Have Taken Down My Post*, FACEBOOK, https://www.facebook.com/help/2090856331203011?helpref=faq_content [https://perma.cc/FU3E-AT4E].

humans review the content's compliance with a platform's rules.[70]

Much recent content moderation scholarship has been about other people involved in content moderation—namely, Facebook's Oversight Board.[71] The Board is supposed to rationalize and publicize some content moderation decisions, generating trust and confidence.[72] Notably, the Board's members are different than most others involved in the content moderation process. Designers, content creators, commercial content moderators, and other users have no particular expertise in classifying a piece of content as violative of platform rules. Presumably, few of them have studied pornography or terrorism or hate speech. They may have read their platforms' community guidelines, but even outsourced content moderators are not hired for their expertise in free speech law.[73] Board members, on the other hand, are supposed to be "experienced at deliberating . . . and familiar with digital content and governance" and have "demonstrated a proficiency in questions of online content moderation."[74] A majority of Board members are law professors with expertise in free speech. In other words, the content moderation that happens at the Board level is supposed to be done by those people Facebook considers experts; at the flagging, appealing, and reviewing level, where far more decisions are made, content moderators are laypersons, with no particular expertise beyond whatever training Facebook offers or their own interpretation of the platform's community guidelines. This expertise (or lack thereof) plays a central role in both the sexual content moderation and anti-vice policing assemblages.

B. Putting Moderation Back Together

Law and technology scholarship has independently and sometimes implicitly recognized that content moderation is a multifaceted institution comprising different social actors. The last section reorganized the literature to see content moderation as a product of different elements—legal contexts, discourses, rules, code, AI, and people—coming together to exert control over content in a way that none of them could independently. Indeed, as Jameel Jaffer suggested, content moderation on Facebook is a combination of an "interface, algorithms, and

70. Klonick, *Governors*, *supra* note 7, at 1638–41.

71. Kate Klonick, *The Facebook Oversight Board*, 129 YALE L.J. 2418 (2020).

72. *Id.* at 2449–50.

73. Adrien Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, WIRED (Oct. 23, 2014, 6:30 AM), https://www.wired.com/2014/10/content-moderation/ [https://perma.cc/M8EN-H49U] [hereinafter Chen, *Laborers*].

74. OVERSIGHT BD., https://oversightboard.com/meet-the-board/ [https://perma.cc/S4WQ-5ANZ].

policies . . . that determine[] which speech is amplified and which is suppressed."[75] Rules without design constraints are easily flouted; AI filters will misunderstand things humans could contextualize; the black box of code cannot maintain a platform's ethos without transparent rules to which individuals can refer when uploading content.[76] In this way, content moderation is more than the sum of its parts. As the above review suggests, content moderation is an assemblage. It "make[s] something happen" through both the actions and affordances of its elements, but also from the assemblage as a whole.[77]

And yet, although the current literature has studied most of content moderation's elements independently, most scholars fall back on analyzing the institution broadly.[78] Scholars cite a wide variety of content moderation examples—videos of mass shootings,[79] gratuitous gore violence,[80] slanderous cartoons,[81] predatory advertisements for masks during the COVID-19 pandemic,[82] images of child sexual abuse,[83] content calling for election interference,[84] incidents of gendered harassment,[85] and terrorism,[86] among others—to support conclusions about content moderation generally as if it is a singular institution doing everything at once. Although important and generative, that scholarship is incomplete. There are different types of content moderation, each of which is its own assemblage of social forces that might operate with different ideas, influences, rules, technology, and people. Only a few sociolegal scholars have focused exclusively on sexual content moderation,[87] and even fewer have done so through the lens of queer sexuality.[88]

---

75. Jameel Jaffer, *Facebook and Free Speech Are Different Things*, KNIGHT FIRST AMEND. INST. (Oct. 24, 2019), https://knightcolumbia.org/content/facebook-and-free-speech-are-different-things [https://perma.cc/L3A6-PXG3].

76. GILLESPIE, *supra* note 7, at 47.

77. JANE BENNETT, VIBRANT MATTER: A POLITICAL ECOLOGY OF THINGS 24 (2010).

78. *E.g.*, Douek, *Systems*, *supra* note 13, at manuscript 9–39.

79. Bloch-Wehba, *supra* note 7, at 42.

80. *Id.*; Klonick, *Governors*, *supra* note 7, at 1619–20, 1641.

81. Klonick, *Governors*, *supra* note 7, at 1623–24.

82. Douek, *Governing*, *supra* note 6, at 808.

83. Bloch-Wehba, *supra* note 7, at 57.

84. Douek, *Governing*, *supra* note 6, at 779.

85. *Id.*

86. Bloch-Wehba, *supra* note 7, at 58.

87. *E.g.*, Gerrard & Thornham, *supra* note 33; Rena Bivens & Oliver Haimson, *Baking Gender Into Social Media Design: How Platforms Shape Categories for Users and Advertisers*, 2 SOC. MEDIA & SOC'Y 1 (2016).

88. *See* Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet As It*

What is more, most content moderation scholarship, not to mention popular commentary, focuses on the system's mistakes: for example, the pictures of breastfeeding mothers that should not have been taken down,[89] the deplatforming of anti-racist skinheads instead of neo-Nazis,[90] the failure to remove a violent militia page,[91] or the inappropriate removal of a picture of "Napalm Girl," a Pulitzer Prize-winning photograph from the Vietnam War.[92] Understanding why those mistakes occur and how to mitigate them is critical, particularly when errors chill the speech of traditionally marginalized groups. But, as Sarah Roberts notes, focusing too much on mistakes takes attention away from systemic interrogation of content moderation's machinery and the effects of its design.[93] In other words, if every problematic content moderation decision is a mistake, especially when platforms claim it was ex post, we legitimize the underlying system as a whole by conceptualizing its effects as aberrations.

The next Part begins to fill those gaps by showing how the assemblage of social forces that come together to exert power over public expressions of queer sexuality online resembles anti-vice policing. This assemblage of law, discourses, technology, and people "smothered" public expressions of queer sexuality in the middle of the last century, and the same forces are operating online today.[94] This analogy shows just how much a queer perspective on content moderation differs from the conventional wisdom. From this analogy, we can develop new paradigms for understanding and solving content moderation's systemic faults.

II. A HISTORICAL MODEL FOR SEXUAL CONTENT MODERATION

Reframing the scholarly literature to surface the assemblage of actors that exert power over user-generated content challenges how we think about content moderation. This creates new opportunities for insight about

---

Is (And As It Should Be), 118 MICH. L. REV. 1073 (2020); Yoel Roth, *"No Overly Suggestive Photos of Any Kind": Content Management and the Policing of Self in Gay Digital Communities*, 8 COMMC'N, CULTURE & CRITIQUE 414 (2015); Bonnie Ruberg, *"Obscene, Pornographic, or Otherwise Objectionable": Biased Definitions of Sexual Content in Video Game Live Streaming*, 23 NEW MEDIA & SOC'Y 1681 (2021).

89. Douek, *Governing*, *supra* note 6, at 774.

90. *Id.* at 808.

91. *Id.* at 47–48.

92. Roberts, *Detritus*, *supra* note 7.

93. *Id.*

94. William N. Eskridge, Jr., *Challenging the Apartheid of the Closet: Establishing Conditions for Lesbian and Gay Intimacy,* Nomos*, and Citizenship, 1961–1981*, 25 HOFSTRA L. REV. 817, 819 (1997) [hereinafter Eskridge, *Challenging*].

content moderation's structure, applications, and weaknesses. This Article's next goal is to explore the machinery of one of those assemblages—sexual content moderation—through an explicitly queer lens. This is long overdue. Despite several high-profile incidents of queer censorship online,[95] there has been no systematic exploration in the legal literature of the origins of and manner in which content moderation exerts control over queer content.[96]

This Part argues that sexual content moderation is an assemblage comprising similar forces and operating in ways similar to anti-vice policing in urban centers in the United States from the 1930s through the 1960s. Reacting to similar contexts and discourses, both platforms and municipalities use similarly vague rules, with similar rationales and underlying assumptions, to asymmetrically crack down on queer behavior in public, denying queer people due process and contributing to similar outcomes: the systematic erasure of queer, nonnormative behavior from public spaces. The extent to which sexual content moderation recreates the heteronormative assemblage of anti-vice policing has profound implications for the way we think about content moderation, for the role of technology in society, for the role of expertise in constructions of knowledge and power, and for debates about platform regulation.

This Part begins by paralleling the social contexts and rationales underlying both assemblages. It then turns to their similar rules and mechanics. It concludes by demonstrating the comparison's durability while acknowledging important limitations.

A. Legal Contexts, Discourses, and Values

Anti-vice policing is part of almost every retelling of queer life in the United States. The narrative has allowed legal and political historians to

---

95. *E.g.*, Sarah Perez, *Tumblr Says It Fixed the 'Safe Mode' Glitch that Hid Innocent Posts, Including LGBTQ+ Content*, TECH CRUNCH (June 24, 2017, 9:45 AM), https://techcrunch.com/2017/06/24/tumblr-says-it-fixed-the-safe-mode-glitch-that-hid-innocent-posts-including-lgbtq-content/ [https://perma.cc/REX9-QKYD]; Elle Hunt, *LGBT Community Anger Over YouTube Restrictions Which Make Their Videos Invisible*, GUARDIAN (Mar. 19, 2017), https://www.theguardian.com/technology/2017/mar/20/lgbt-community-anger-over-youtube-restrictions-which-make-their-videos-invisible [https://perma.cc/Q8A6-45SU].

96. That said, scholars in other fields are considering the systemic anti-queer effects of content moderation. In addition to the scholars cited throughout this Article, please see SUSANNA PAASONEN, KYLIE JARRETT & BEN LIGHT, NSFW: SEX, HUMOR, AND RISK IN SOCIAL MEDIA (2019); Thiago Dias Oliva, Dennys Marcelo Antonialli & Alessandra Gomes, *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, 25 SEXUALITY & CULTURE 700 (2021). *See also*, *e.g.*, Danielle Keats Citron, *Cyber Civil Rights*, 89 B.U. L. REV. 61 (2009) [hereinafter Citron, *Civil Rights*]. Citron's research incorporates experiences of queer people, particularly those victimized by nonconsensual pornography and other forms of online harassment. *E.g.*, Citron, *Sexual Privacy*, *supra* note 7. Her work is foundational and complementary.

describe the queer community's political awakening, its resistance to oppression on the ground, and its use of the law on the books to push back.[97] Others have used the backdrop of morality policing to teach lessons about criminal law, the limits of policing in general, and the extent to which the legal system has been a site of contestation over the defining features of queerness.[98] Queer people's confrontations with morality law can also teach us about content moderation, starting with its origins and rationalizing discourses.

1. Repressive Origins

Anti-vice policing in the twentieth century arose, evolved, and reached its apex at times of overlapping and profound cultural shifts: from Prohibition to the Depression, from mass mobilization for World War II to the post-war return to the cities, from growing queer populations to moral panics, and from the promise of liberal experimentation to the retrenchment of conservative values.[99] Content moderation evolved during another cultural shift from promise to peril, one in which early cyber-utopianism gave way to the reality of hate, predation, and corporate control online. In both cases, liberality was always unsustainable; for many queer people, both the promise of and repression against queer behavior were real.

Before the Depression, queer life integrated into the everyday life of cities like New York and Los Angeles.[100] Prohibition created a culture of flouting conventions, including sexual ones. Speakeasies were "liminal spaces in which visitors were encouraged to disregard some of the social injunctions that normally constrained their behavior" in their home neighborhoods.[101] In this permissive environment, "subversive" queer culture became part of urban culture.[102] Prohibition's repeal and the subsequent reimposition of municipal regulatory control over liquor sales "inaugurated a more pervasive and more effective regime of surveillance of control" over urban nightlife.[103] Suddenly, the police were everywhere

97. In addition to the texts already cited, please see ERIC CERVINI, THE DEVIANT'S WAR (2020); GEORGE CHAUNCEY, GAY NEW YORK (1994); JOHN D'EMILIO, SEXUAL POLITICS, SEXUAL COMMUNITIES (1983); NAN ALAMILLA BOYD, WIDE OPEN TOWN (2003); LILLIAN FADERMAN & STUART TIMMONS, GAY L.A. (2006); ELIZABETH LAPOVSKY KENNEDY & MADELINE D. DAVIS, BOOTS OF LEATHER, SLIPPERS OF GOLD (2003).

98. *See* LVOVSKY, *supra* note 9.

99. *Id.* at 5.

100. CHAUNCEY, *supra* note 97, at 4.

101. *Id.* at 305–08.

102. *Id.* at 4, 301.

103. *Id.* at 335.

in New York's West Village and Times Square, eager to use their power to reintroduce the social boundaries that existed before swing dance, flappers, pansies, and drag balls.[104]

Even more profound shifts took place later, catalyzing the worst of anti-vice policing in the post-World War II era.[105] Soldiers returning from war increasingly settled in urban areas; the rise of queer neighborhoods, bars, and cruising grounds struck many of them as inappropriate.[106] A rash of highly publicized sex crimes caused a nationwide panic about the morality of urban life.[107] And the Lavender Scare, the opportunistic campaign to remove queer people from the federal government, framed queer men as dangerous security risks.[108] These social forces gave added purpose to anti-vice policing at a time when society as a whole was taking a keen interest in heteronormative domesticity.

This cultural shift from the promise of liberality to the perils of repression parallels the evolving understanding of the Internet as a locus of freedom, expression, and individuality. Originally, commentators thought the Internet would be an ideal space for free expression. The Internet evangelist John Perry Barlow called it "a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity,"[109] not unlike John Locke's description of the state of nature.[110] Early law and technology scholars saw cyberspace as a space where anyone who wanted "a variety of topics or views will easily be able to get them."[111] There would always be more speakers, more listeners, more information, and more opportunity, all leading to the ultimate realization of a true "marketplace of ideas."[112]

---

104. *Id.*

105. LVOVSKY, *supra* note 9, at 4–6.

106. ALLAN BÉRUBÉ, COMING OUT UNDER FIRE 67–127 (1990).

107. *See* Andrea Friedman, *Sadists and Sissies: Anti-Pornography Campaigns in Cold War America*, 15 GENDER & HIST. 201, 214–17 (2003).

108. DAVID JOHNSON, THE LAVENDER SCARE (2009). Because of the social constraints of the closet and the profound stigma associated with homosexuality in the middle of the twentieth century in the U.S., perpetrators of the "lavender scare" presumed that gay government workers would be easy targets for blackmail by communist or foreign powers. *Id.* at 5–15.

109. John Perry Barlow, *A Declaration of the Independence of Cyberspace*, ELEC. FRONTIER FOUND. (Feb. 8, 1996), https://projects.eff.org/~barlow/Declaration-Final.html [https://perma.cc/D5UR-V8XN].

110. Mary Anne Franks, *Unwilling Avatars: Idealism and Discrimination in Cyberspace*, 20 COLUM. J. GENDER & L. 224, 234–35 (2011) [hereinafter Franks, *Unwilling Avatars*].

111. Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L.J. 1805, 1834 (1995); *see also* Kathleen M. Sullivan, *First Amendment Intermediaries in the Age of Cyberspace*, 45 UCLA L. REV. 1653, 1670 (1998).

112. Franks, *Unwilling Avatars*, *supra* note 110, at 234–37.

In cyberspace, the law was supposed to leave people alone, allowing them to express themselves authentically without state surveillance or repression. Therefore, it should come as no surprise that some scholars also saw promise for queer people in the early Internet. Nearly a decade ago, the Gay, Lesbian, and Straight Education Network (GLSEN) found that social media allowed queer youth to "find greater peer support, access . . . health information and [find] opportunities to be civically engaged."[113] Studies showed that the Internet provided an outlet for those living closeted lives.[114] Noting that functional pseudonymity gave queer people the chance to meet, explore their sexuality, and support activism outside the reach of hostile communities, Edward Stein argued that attempts to regulate online speech threatened real and lasting harm to queer people who depended on the protection of the closet.[115] The law's long history of complicity in queer oppression also made the prospect of a space without law enforcement particularly attractive to queer people.[116]

This idealism ran up against cultural shifts calling for tighter regulation of online expression. Like early anti-vice enforcement, which began during the conservative retrenchment after Prohibition, the earliest forms of content moderation began during the culture wars of the 1990s and 2000s, when conservative politicians leveraged white, middle-class unease with abortion, homosexuality, and pornography to gain electoral advantage and define American identity in line with patriarchal values.[117] Those culture wars had direct effects on Internet expression. Bipartisan majorities in Congress passed and President Clinton signed the Communications Decency Act in 1996—just a few years after the World Wide Web became generally accessible to most Americans—to protect minors from incident material online.[118] The Family Enforcement Act, which was introduced at the same time, called for enhanced punishment

113. GAY, LESBIAN & STRAIGHT EDUCATION NETWORK, OUT ONLINE: THE EXPERIENCES OF LESBIAN, GAY, BISEXUAL AND TRANSGENDER YOUTH ON THE INTERNET (2013), http://www.glsen.org/press/study-finds-lgbt-youth-face-greater-harassment-online [https://perma.cc/E8Y5-3RUZ].

114. Ronny Tikkanen & Michael W. Ross, *Technological Tearoom Trade: Characteristics of Swedish Men Visiting Gay Internet Chat Rooms*, 15 AIDS EDUC. & PREVENTION 122, 122 (2003); Jonathan Alexander, *Queer Webs: Representations of LGBT People and Communities on the World Wide Web*, 7 INT'L J. SEXUALITY & GENDER STUD. 77, 78–79 (2002).

115. Edward Stein, *Queers Anonymous: Lesbians, Gay Men, Free Speech, and Cyberspace*, 38 HARV. C.R.-C.L. L. REV. 159, 161–63 (2003).

116. KEVIN NADAL, QUEERING LAW AND ORDER (2020).

117. *See* THOMAS FRANK, WHAT'S THE MATTER WITH KANSAS (2004); ANDREW HARTON, A WAR FOR THE SOUL OF AMERICA 1–7 (2d ed. 2019).

118. Telecommunications Act of 1996, Pub. L. No. 104–104, §§ 501–09, 110 Stat. 56, 133–39 (1996). It is true that the Supreme Court declared the Communications Decency Act (except for section 230) unconstitutional in *Reno v. ACLU*, 521 U.S. 844 (1997).

if computers were used to share child pornography.[119] Similar legislation was also introduced in the states.[120] Universities also blocked access to pornographic material after a (now-debunked) study showed a significant amount of pornography on the early Internet.[121]

To be clear, these social contexts are not identical; the anti-queer panic of the 1950s does not match, either in its severity or its discourses, the anti-queer, anti-porn panics of the 1990s. My argument is more modest. It was the retrenchment of conservative values after Prohibition, and then again after World War II, that formed the background context for cities like New York and Los Angeles to systematically regulate queer sexual behavior. Similarly, it was during the conservative retrenchment of the culture wars of the 1990s and 2000s that platforms started moderating sexual content and writing rules to routinize it. Both phenomena arose in cultural moments, unique to their times, where rule makers were thinking much more about regulating public sexual expression.

2. *Discourses Rationalizing Repression*

As they turned their attention to sex, policymakers in both eras channeled similar anxieties about pornography, predation, and lawlessness into rationales about satisfying business, protecting children from exposure to indecent material, and reasserting their control over spaces after threats to take it away. This section demonstrates the prevalence of each rationale in turn.

a. *Economic Rationales*

Both mayors and social media companies thought public expressions of queerness were bad for business. New York Mayor Fiorello LaGuardia instructed the police to work with the State Liquor Authority (SLA), which regulated all New York establishments that served alcohol after Prohibition, to shut down all the queer and queer-friendly bars in the City

---

119. H.R. 11, 104th Cong., 1st Sess. § 301 (1995); David Rosenbaum, *The 1994 Campaign: The Republicans*, N.Y. TIMES (Nov. 1, 1994), https://www.nytimes.com/1994/11/01/us/ 1994-campaign-republicans-it-s-economy-again-democrats-attack-contract-with.html [https://perma.cc/SG7P-98D9] (describing the Family Reinforcement Act as part of the Republican Contract with America).

120. *See* Carlin Meyer, *Reclaiming Sex from the Pornographers: Cybersexual Possibilities*, 83 GEO. L.J. 1969, 1971 n.20 (1995) (citing N.Y. Assem. B. 3967, 218th Gen. Ass., 1st Sess. (1995) as an example of state responses).

121. Marty Rimm, *Marketing Pornography on the Information Superhighway*, 83 GEO. L.J. 1849 (1995). The study was debunked. Donna Hoffman & Thomas Novak, *A Detailed Analysis of the Conceptual, Logical, and Methodological Flaws of the Article: "Marketing Pornography on the Information Superhighway"*, HE SAID/SHE SAID (July 2, 1995), http://alumni.media.mit.edu/~rhodes/Cyberporn/hn.on.rimm.html [https://perma.cc/MW7W-GSJW].

in anticipation of the 1939 World's Fair and its attendant influx of people, tourism, and money.[122] As the historian Lewis Erenberg argued, the repeal of Prohibition and the subsequent passage of state laws to regulate the consumption of alcohol were essential for cities to lure back "reputable" businesses that had been decimated or scared off by Prohibition's lawlessness.[123] Closing down businesses that catered to a queer clientele opened spaces for those supposedly "reputable" businesses and the recreation of a "sanitized" nightlife that would attract a crowd far wealthier than the working class, immigrant, and queer men who dominated queer nightlife during the 1920s and early 1930s.[124] The authorities assumed that no legitimate business would want to set up shop around a queer crowd.

Similar assumptions and rationales pervade justifications for sexual content moderation. When YouTube demonetized a series of lesbian sex education videos, the only explanation it provided to the content creator was that the series was "not suitable for all advertisers."[125] Tumblr, which had failed to meet two years of advertising goals set by its parent company Yahoo, restricted queer sexual content because Yahoo wanted to attract advertisers that had been scared off by the large percentage of adult content on the platform.[126] In 2018, Patreon suspended many of its adult content creators because the financial transaction platforms Stripe and PayPal refused to be associated with sex workers.[127] As Niva Elkin-Koren has argued, platforms may deny it, but content moderation is a way for platforms to ensure that their services "align[] . . . with the interests of advertisers."[128]

b. Reasserting Institutional Control

Regulating sexual content is not only about money. It is also about

---

122. CHAUNCEY, *supra* note 97, at 340.

123. Lewis A. Erenberg, *From New York to Middletown: Repeal and the Legitimization of Nightlife in the Great Depression*, 38 AM. Q. 761, 762–66 (1986).

124. CHAUNCEY, *supra* note 97, at 336–38.

125. Priddy, *supra* note 2.

126. Clare Southerton, Daniel Marshall, Peter Aggleton, Mary Lou Rasmussen & Rob Cover, *Restricted Modes: Social Media, Content Classification, and LGBTQ Sexual Citizenship*, 23 NEW MEDIA & SOC'Y 920, 925 (2021); Kaitlyn Tiffany, *When Tumblr Bans Porn, Who Loses?*, VOX (Dec. 4, 2018, 5:00 PM), https://www.vox.com/the-goods/2018/12/4/18126112/tumblr-porn-ban-verizon-ad-goals-sex-work-fandom [https://perma.cc/H7Y7-3HHD]; Perez, *supra* note 95.

127. Samantha Cole, *Patreon Is Suspending Adult Content Creators Because of Its Payment Partners*, VICE: MOTHERBOARD (June 28, 2018, 2:00 PM), https://www.vice.com/en/article/vbqwwj/patreon-suspension-of-adult-content-creators [https://perma.cc/S8UF-RN9L].

128. Elkin-Koren, *supra* note 32, at 5.

power. In both the anti-vice policing and sexual content moderation contexts, those in power had institutional rationales for reasserting their control over spaces in which that control had been challenged. Both sets of those challenges came from public law: For states and municipalities, Prohibition and national mobilization for World War II undermined government control over nightlife; for platforms, it was the threat of imposing traditional intermediary liability law on internet platforms.

Prohibition had the perverse effect of eroding police and municipal control over urban nightlife. Police were not wholly absent from urban nightlife, but as George Chauncey argued, the Prohibition Era "obliterated" the "boundaries of acceptable sociability" because its culture of lawlessness removed from urban enclaves the institutions that policed those boundaries.[129] The nationwide mobilization for World War II upended traditional gender roles, brought queer people from all over the country into close proximity with each other, and refocused national and municipal attention on something other than queer life.[130] Waves of queer soldiers, emboldened by queer friendships they made in the trenches, returned to the cities and continued to build queer community.[131] Therefore, if Prohibition's repeal was a chance for local governments to reassert social boundaries under their terms, the end of World War II gave cities new reasons to redouble their efforts.[132]

In the anti-vice context, then, municipalities were reasserting control over spaces another arm of the state had taken away. Content moderation may have evolved during a more complex contest for control over online expression between government and private platforms, but the contexts are similar in that both anti-vice enforcement and content moderation rules were attempts by the powerful to reassert control that had been taken away.

In *Stratton Oakmont, Inc. v. Prodigy Services Co.*,[133] a New York trial court threatened to impose traditional publisher liability rules on Prodigy, an early online service provider that held itself out as "family oriented" and willing to remove profanity and hate.[134] When Prodigy failed to remove defamatory comments about an investment firm from one of the

---

129. CHAUNCEY, *supra* note 97, at 334–35. A related cultural phenomenon occurred after World War II. As Anna Lvovsky argued, former soldiers who settled in urban centers to find jobs after the War found queer enclaves, bars, and cruising grounds a "blight on the orderly city." LVOVSKY, *supra* note 9, at 5.

130. LVOVSKY, *supra* note 9, at 5.

131. *Id.* at 102.

132. *Id.* at 5, 9, 20, 101, 147.

133. No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

134. *Id.* at *2.

platform's bulletin boards, the firm sued Prodigy on the theory that the platform was liable as a publisher of the defamation because its filtering policies determined what would be allowed on its platform.[135] The trial court held Prodigy liable for two-hundred million dollars,[136] reasoning that if Prodigy actively asserts control over the content, it is responsible for that content.[137] Therefore, *Stratton Oakmont* represented a threat to platforms' control over their spaces. In *Stratton Oakmont*, the state was trying to step in, wresting control away from online platforms in the same way that it uses tort law to limit the editorial discretion of newspapers. Communications Decency Act section 230, which immunized platforms from tort liability associated with any user content and, therefore, gave platforms the breathing space to develop their own content moderation rules, was passed in reaction to that threat.[138] Understood in this way, content moderation rules were attempts to reassert the kind of unlimited control that publisher liability law threatened to take away.

c. Protecting the Children and Other Vulnerable Populations

As noted above, anti-vice police attention to queer behavior in public came at a time of cultural anxiety about upended social and sexual norms. One discursive weapon used to justify to reassert those traditional norms was the protection of children.[139] Bill Eskridge's exhaustive studies of law relating to gender and sexual nonconformity demonstrate this. The law reflected the archetype of the queer man as predatory, seeking children to harm and convert.[140] Anti-vice societies—civic groups that supported state and municipal crackdowns on prostitution and all forms of what they called "moral degeneracy"—focused on protecting children as well. One of the most famous of those clubs in New York was the Society for the Prevention of Cruelty to Children, the goal of which was to protect adolescents from "child molesters" through aggressive enforcement of

135. *Id.*

136. DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 169 (2014) [hereinafter CITRON, HATE CRIMES].

137. *Stratton Oakmont*, 1995 WL 323710, at *3.

138. Citron & Wittes, *supra* note 37, at 404–06.

139. Notably, protecting children was not the only moralistic justification for anti-vice policing. Many police and policymakers were also responding to broader backlashes against homosexuality, including the Lavender Scare crusade against queer employees of the federal government, and panics about sex crimes and sexual predators. LVOVSKY, *supra* note 9, at 5; JOHNSON, *supra* note 108, at 73.

140. ESKRIDGE, GAYLAW, *supra* note 16, at 4; JOEY L. MOGUL, ANDREA J. RITCHIE & KAY WHITLOCK, QUEER (IN)JUSTICE 31–34, 77–90 (2011); William N. Eskridge, Jr., *Law and the Construction of the Closet: American Regulation of Same-Sex Intimacy, 1880–1946*, 82 IOWA L. REV. 1007, 1059 (1997).

anti-sodomy laws.[141] Laws prohibiting public lewdness, indecency, vagrancy, and disorderly conduct "aimed at preventing the corruption of children" by queer men.[142] Crackdowns on same-sex activity in the military after World War I leaned hard on the youth angle as well. "The people of the United States," read one salient report, "are entitled to the assurance that thereafter no *boy* who enlists in the Navy will be consigned to a career of vice."[143] A Senate committee investigating the matter was concerned that "*boys*" were joining the military to be "the prey of every degenerate or sexual pervert."[144]

Protecting children also motivates sexual content moderation. Children's sexual innocence was one rationale for the CDA, which sought to restrict children's access to indecent material online.[145] Congress wrote the CDA out of a moralistic concern about pornography, in general.[146] It was also an animating force for the Online Family Empowerment Act.[147] Platforms use similar discourses when they make their sexual content moderation more robust. Instagram removed sex educational content because the platform wants "content to be appropriate for . . . our youngest members."[148] YouTube hid queer content in Restricted Mode after stating that "[c]hild safety has been and remains" the company's "#1 priority."[149] TikTok justified its increasingly robust filtering of sexual content as a way "to further strengthen our safeguards and introduce new measures to protect young people on the app."[150] The result was an entire

141. ESKRIDGE, GAYLAW, *supra* note 16, at 23.

142. *Id*. at 27.

143. *Id*. at 38 (emphasis added).

144. *Id*. (citing Alleged Immoral Conditions at Newport (R.I.) Naval Training Station, Rep. of the Comm. on Naval Affairs, U.S. Senate, 67th Cong., 1st sess. (1921)) (emphasis added).

145. S. Rep. No. 104-23, at 9, 59 (noting that law was adopted to, among other things, "protect families from uninvited cable programming which is unsuitable for children"); *id*. at 59 ("The information superhighway should be safe for families and children.").

146. *Id.* at 59 ("The Committee has been troubled by an increasing number of published reports of inappropriate uses of telecommunications technologies to transmit pornography.").

147. H.R. Rep. No. 104-223, at *14 ("Congress finds the following: . . . (4) remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material.").

148. Abigail Moss, *'Such a Backwards Step': Instagram Is Now Censoring Sex Education Accounts*, VICE (Jan. 8, 2021, 6:56 AM), https://www.vice.com/en/article/y3g58m/instagram-rules-censoring-sex-educators [https://perma.cc/ZYR5-CYDQ].

149. Matthew S. Schwartz, *Advertisers Abandon YouTube Over Concerns That Pedophiles Lurk in Comments Section*, NPR (Feb. 22, 2019, 8:53 AM), https://www.npr.org/2019/02/22/696949013/advertisers-abandon-youtube-over-concerns-that-pedophiles-lurk-in-comments-secti [https://perma.cc/FTP7-2FTX].

150. Justin Wise, *Feds Investigating Allegations TikTok Failed to Protect Children's Privacy: Report*, HILL (July 7, 2020, 11:27 PM), https://thehill.com/policy/technology/506323-feds-investigating-allegations-tiktok-failed-to-protect-childrens-privacy [https://perma.cc/PS6T-7URP].

hashtag—#wrongsideoftiktok—that includes, among other things, content algorithmically labeled as sexual in any way and slowed, taken down, or entirely removed under the guise of protecting children.[151]

There are, of course, other factors contributing to the evolution of sexual content moderation. For instance, thanks to the work of scholars like Danielle Citron and Mary Anne Franks and victims' rights lawyers like Carrie Goldberg, platforms have focused more attention on protecting women and members of other marginalized populations from online harassment, sexual exploitation, and nonconsensual pornography.[152] Platforms have developed ways to spot and report cyber sexual abuse, even if the process is imperfect and frustrating.[153] Some of these changes come from a refreshing victim-centered approach.[154] Protecting the cyber civil rights of women is a legitimate goal, as is protecting children from being force-fed explicit or obscene content. I am not suggesting otherwise. My narrower point is that, like anti-vice policing, sexual content moderation has been rationalized as a means to protect vulnerable populations from exploitation. Undoubtedly, the discourse of protecting children can and has been weaponized as a front for anti-queer bigotry.[155] But even if we credit platforms with genuine concern for preventing sexual exploitation, the next section demonstrates how, like anti-vice policing, the machinery of sexual content moderation's vague rules, technologies, and people ends up disproportionately silencing queer expression as well.

---

151. *See* Morgan Sung, *The Stark Divide Between 'Straight TikTok' and 'Alt TikTok'*, MASHABLE (June 21, 2020), https://mashable.com/article/alt-tiktok-straight-tiktok-queer-punk [https://perma.cc/RAU6-V8YD].

152. *See* CITRON, HATE CRIMES, *supra* note 136; Citron, *Civil Rights*, *supra* note 96; Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251 (2017); Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014). Carrie Goldberg runs a pioneering practice representing victims of online sexual abuse. *See* C.A. GOLDBERG, https://www.cagoldberglaw.com/ [https://perma.cc/8YJQ-ZV5H].

153. CITRON, HATE CRIMES, *supra* note 136, at 237–39.

154. Olivia Solon, *Inside Facebook's Efforts to Stop Revenge Porn Before It Spreads*, NBC NEWS (Nov. 19, 2019, 8:15 AM), https://www.nbcnews.com/tech/social-media/inside-facebook-s-efforts-stop-revenge-porn-it-spreads-n1083631 [https://perma.cc/85RZ-VNAQ].

155. *See* DUDLEY CLENDINEN & ADAM NAGOURNEY, OUT FOR GOOD 292 (2013). The primary thrust of conservatives' arguments against repealing anti-sodomy laws and permitting same-sex marriage were based on perceived harms to children. *See* Courtney G. Joslin, *Searching for Harm: Same-Sex Marriage and the Well-Being of Children*, 46 HARV. C.R.-C.L. L. REV. 81, 85–90, 93–94 (2011). More recently, bans on teaching critical race theory and the resurrection of "Don't Say Gay" or "No Promo Homo" laws in Florida and elsewhere have also been rationalized as necessary to protect children. *See* Mark Joseph Stern, *How the War on Critical Race Theory Revived Anti-Gay Activism in Schools*, SLATE (Mar. 2, 2022, 2:50 PM), https://slate.com/news-and-politics/2022/03/critical-race-theory-dont-say-gay-florida-lgbtq.html [https://perma.cc/4SWZ-MHBF] ("Whenever there is a moral panic involving children, homophobes see an opportunity.").

B. *The Assemblages' Mechanics*

Armed with various rationales for exerting control over public spaces, both states enforcing anti-vice laws and platforms moderating sexual content created a system of vague rules that never mentioned queer people, but necessarily resulted in disproportionate burdens on sexually nonnormative populations.[156] And both regimes rely implicitly on similar sets of assumptions about identifying offending behavior.

1. *Vague Rules*

Most of the laws that police used to harass, arrest, and imprison queer people in the middle of the twentieth century were vague.[157] Police in New York arrested queer men for "disorderly conduct" for a litany of behaviors, from "using vile language" to "not behaving," from walking together on the street to speaking into each other's ears.[158] Rule 5, New Jersey's disorderly conduct provision, prohibited bars from operating "in such a manner as to become a nuisance."[159] The District of Columbia prohibited queer behavior through laws that criminalized "unlawful assembly, profane, and indecent language," as well as "indecent exposure."[160] In cities across the country, mostly gay men were arrested for "degeneracy" or "lewd" or "indecent" conduct.[161] Queer men cruising in public parks for consensual companionship were arrested for "frequent[ing] or loiter[ing] about any public place."[162] Public nuisance laws, as well as "sex degeneracy" and "sex perversion" laws, allowed

156. CHAUNCEY, *supra* note 97, at 2.

157. It is true that until 1961, every state criminalized sodomy, which they variously defined as any oral or anal sex. But because sodomy was difficult to prove, police relied on vague vice laws. LVOVSKY, *supra* note 9, at 104, 195–97.

158. CHAUNCEY, *supra* note 97, at 170, 172, 174, 185, 338. Indeed, most queer men arrested in New York during this time were arrested for "disorderly conduct." *See* CHAUNCEY, *supra* note 97, at 185; ESKRIDGE, GAYLAW, *supra* note 16, at 26–33; Eskridge, *Challenging*, *supra* note 94, at 851, 855, 860, 871–72, 901.

159. LVOVSKY, *supra* note 9, at 51.

160. William N. Eskridge, Jr., *Privacy Jurisprudence and the Apartheid of the Closet, 1946–1961*, 24 FLA. ST. L. REV. 703, 711 (1997) [hereinafter Eskridge, *Privacy*] (citing D.C. CODE ANN. § 22-2701 (West 1940) (inviting for purposes of public prostitution), § 22-1107 (unlawful assembly, profane and indecent language), and § 22-1112 (indecent exposure)).

161. CHAUNCEY, *supra* note 97, at 173, 183, 194; ESKRIDGE, GAYLAW, *supra* note 16, at 26–33; Eskridge, *Challenging*, *supra* note 94, at 837. The historian Nayan Shah reports that police often used these vague laws not just to police queer conduct in general, but particularly consensual sexual encounters between older immigrant men and younger, white men. Nayan Shah, *Between "Oriental Depravity" and "Natural Degenerates": Spatial Borderlands and the Making of Ordinary Americans*, 57 AM. Q. 703 (2005).

162. CHAUNCEY, *supra* note 97, at 185; MOGUL ET AL., *supra* note 140, at 46–47 (queer men were arrested during raids on bars for "loitering inside a building").

police to arrest men who gathered together on stoops or sang showtunes in public.[163] Such laws even formed the basis for pretextual arrests.[164] Men in bars could be arrested under disorderly conduct laws simply for "appear[ing] to be homosexuals."[165] Together, these vague laws formed the basis of an oppressive regime colloquially known as "walking while gay." A version of this regime lives on today as police have used some of the same laws still on the books to harass, arrest, and imprison transgender individuals, sex workers, and queer people of color on the streets.[166]

Few of these laws were specific. As Professor Lvovsky notes, there was no single or clear definition of any of the terms used. The state knew it wanted to marginalize queer sexual conduct, but it had no single "paradigm of so-called deviance."[167] It left room for multiple views held by the various people and departments in charge of enforcement.[168] In the end, although there was no clear definition for what made a person or a bar "disorderly," police knew it when they saw it.[169]

Many sexual content moderation rules are just as vague as the disorderly conduct statutes that policed queer behavior in the last century.[170] Although its rules on "intercourse" and "sexual exploitation" are relatively specific, Facebook's Community Standards prohibit "sexual activity" but provide a non-exhaustive list of examples that covers only a handful of situations, from "erections" to "stimulating genitalia."[171] The rules even include restrictions on "implied stimulation," but they do not

---

163. CHAUNCEY, *supra* note 97, at 313, 352.

164. As Lvovsky notes, "a rash of publicized sex crimes spawned a panic about degeneracy in the nation's cities." LVOVSKY, *supra* note 9, at 4–5. According to some historians, these "sex crimes panics" were excuses for police to arrest allegedly queer men and force them to reveal the names of other queer men, all of whom were arrested for crimes they never committed. MOGUL ET AL., *supra* note 140, at 33.

165. LVOVSKY, *supra* note 9, at 51 (summarizing cases in which mere appearance or perception of queer clientele could result in arrest and shutting down the bar); *see also* One Eleven Wines & Liquors, Inc. v. Div. of Alcohol Beverage Control, 235 A.2d 12, 14–15 (N.J. 1967).

166. Leonore F. Carpenter & R. Barrett Marshall, *Walking While Trans: Profiling of Transgender Women by Law Enforcement, and the Problem of Proof*, 24 WM. & MARY J. WOMEN & L. 5–6, 14, 16 (2017).

167. LVOVSKY, *supra* note 9, at 18.

168. CHAUNCEY, *supra* note 97, at 171, 250, 334; ESKRIDGE, GAYLAW, *supra* note 16, at 46; HUGH RYAN, WHEN BROOKLYN WAS QUEER 95–96 (2019).

169. Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring) ("I know it when I see it.").

170. There is some disagreement about this in the literature. *Compare* Douek, *Governing*, *supra* note 6, at 773 (referring to "elaborate rules"), *and id*. at 785 (implicitly contrasting contextual evaluation with early content moderation's "adherence to clear rules"), *with* Bloch-Wehba, *supra* note 7, at 84 (referring to the "lack of clarity and apparent arbitrariness of some" content moderation rules).

171. META, *supra* note 23.

explain what constitutes "stimulation," actual or implied.[172] TikTok prohibits "content that explicitly or implicitly depicts" sexual conduct, including "erotic kissing," but does not explain the difference between kissing and "erotic" kissing.[173] YouTube has a rule against "explicit content meant to be sexually gratifying" and provides examples with a warning to users that the "list isn't complete."[174] Evidently, sexual content moderation rules follow a pattern: vagueness.

And that pattern is no accident. Anti-vice authorities purposely chose to use vague morality laws to police public expression of queer sexuality because anything more specific was hard to prove. For sodomy, police needed evidence of specific sexual acts done in private among consenting adults.[175] Short of barging through someone's door or managing to peer through a second or third story window, police were hard-pressed to find hard evidence in most cases.[176] Even homosexual solicitation cases could be muddied when anti-vice police stepped too close to the entrapment line.[177] As a result, police arrested more queer men for the lesser charges of "vagrancy" or "disorderly conduct," neither of which had clear definitions, evidentiary requirements, or significant barriers to prosecution.[178]

Similarly, platforms intentionally leave their content moderations rules vague. As Sarah Roberts explains, platforms refuse to disclose material information about what does and does not constitute a violation of their content moderation rules to stop "unscrupulous users [from] attempting to game the rules."[179] But vagueness is less about deterring user mischief than about maintaining platform power. Since platforms are the only ones that can say what their own rules mean, opacity maximizes platforms' discretion to control their fiefdoms.[180]

---

172. *Id.*; *Adult Sexual Exploitation*, META, https://transparency.fb.com/policies/community-standards/sexual-exploitation-adults/ [https://perma.cc/PWW9-T4MX].

173. TIKTOK, *supra* note 23.

174. YOUTUBE, *supra* note 23.

175. LVOVSKY, *supra* note 9, at 4.

176. In two of the most famous sodomy cases, police had to gain access to private residences to observe sexual activity. *See* Brief of Respondent at 1, Bowers v. Hardwick, 478 U.S. 186 (1986) (No. 85-140), 1986 WL 720442 ("On August 3, 1982, Georgia . . . arrest[ed] 29-year-old Respondent Michael Hardwick in his own bedroom, and charg[ed] him with committing the crime of 'sodomy' with another consenting adult in that very room."); Lawrence v. State, 41 S.W.3d 349, 350 (Tex. App. 2001) (identifying John Lawrence and Tyron Garner engaging in "homosexual conduct" while "investigating a reported 'weapons disturbance'").

177. LVOVSKY, *supra* note 9, at 138.

178. *Id.*

179. ROBERTS, SCREEN, *supra* note 7, at 25, 37.

180. *See* John F. Manning, *Constitutional Structure and Judicial Deference to Agency*

Granted, vagueness in definitions of sexual content is nothing new.[181] The task of defining sexual norms is fraught, culturally situated, and ideological, and I do not suggest otherwise.[182] Rather, the historical analogy between anti-vice policing and sexual content moderation demonstrates how the latter's vague standards, like disorderly conduct and other anti-vice laws, gives ample discretion to human decisionmakers.[183] Decision-makers can refer to operational guidelines with their lists of nonexhaustive examples, but the final decision is discretionary.[184] As the next section describes, implicit in the discretion granted to both police and content moderators is the assumption that anyone can spot inappropriate content, an assumption that further amplifies platform power.

2. Technological and Human Expertise

It is true that mid-century anti-vice policing did not depend for its day-to-day implementation on armies of commercial content moderators from different countries and cultures across the globe with different norms about sex. Although one could argue that queer culture was originally just as foreign and stigmatized to police as it is to many moderators from countries that still ban or criminalize homosexuality, the resemblance between anti-vice policing and sexual content moderation runs deeper.

Both vice squads and social media platforms have a similar enforcement problem: How can anyone know when something violates a vague rule? Other areas of law solve this problem with expertise: for instance, administrative law rests at least in part on the assumption that executive agencies are staffed with experts who are owed deference when writing rules that clarify vague statutory guidelines.[185] Anti-vice policing and sexual content moderation chose the opposite approach. Each system

---

Interpretations of Agency Rules, 96 COLUM. L. REV. 612, 655 (1996) (making a similar argument in the federal agency context).

181. *See* Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring) (stating that trying to define obscenity is trying "to define what may be indefinable"); Miller v. California, 413 U.S. 15, 37–42 (Douglas, J., dissenting) (describing the problems the Supreme Court has faced with defining and applying a rule for obscenity).

182. Ruberg, *supra* note 88, at 1683.

183. LVOVSKY, *supra* note 9, at 7 (describing how anti-vice laws gave great discretion to both police and trial court judges).

184. Roth, *supra* note 88, at 418–420.

185. Expertise is one theory justifying judicial deference to agency decisions in *Chevron U.S.A. Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837 (1984). *See* Yoav Dotan, *Making Consistency Consistent*, 57 ADMIN. L. REV. 995, 1022–23 (2005); Colin S. Diver, *Statutory Interpretation in the Administrative State*, 133 U. PA. L. REV. 549, 596 (1985); Cass R. Sunstein, *Law and Administration After* Chevron, 90 COLUM. L. REV. 2071, 2084 (1990); Frank B. Cross, *Shattering the Fragile Case for Judicial Review of Rulemaking*, 85 VA. L. REV. 1243, 1264 (1999).

implicitly relies on the notion that just about anyone could spot offending behavior.

In addition to using vague laws to constrain individuals' behavior, states and municipalities prohibited bars from becoming "disorderly" by knowingly serving queer customers. But as Lvovsky insightfully argues, by making enforcement contingent on bar owners *knowing* they were serving gay patrons, liquor laws not only required police to be able to identify queer people in bars, they also "forced the states' investigators to prove that [bar owners] had not only welcomed gay customers but also recognized them as such."[186] The burden was often met by suggesting that queer patrons were "unmistakably homosexual" or, implicitly, that there was some common, shared insight about what a queer person looked like.[187] Therefore, the entire regime implicitly depended on the presumption that anyone could identify a queer person on sight.[188]

Undoubtedly, this regime imprinted sexual and gender stereotypes in law. It also implied that anyone could do it. When bars challenged the state's attempt to shut them down for serving queer customers, they often questioned agents' qualifications to identify queer customers.[189] Unwilling to hold themselves out as experts on a lifestyle most of society found abhorrent, police responded that they neither considered themselves experts nor needed to be: "[I know] when I see a pansy or a degenerate," officers claimed in court.[190] Attorneys for the SLA made this explicit, arguing that "[y]ou don't have to be an expert to be able to see a homosexual."[191] That was because everyone thought that all queer people were flamboyant, feminine, and ostentatious, much like they had been portrayed in the media. It was a piece of shared "knowledge" that required no training to deploy.[192] Lvovsky concludes that making inferences of homosexuality were "so very unremarkable" because the police perceived identifying queer people required only "simple deductions that could be corroborated" by anyone.[193]

Like anti-vice policing, sexual content moderation also relies on the presumption of lay expertise in both human and algorithmic moderation. Platforms keep the details of moderator training a secret. But we know that commercial content moderators usually find their jobs through

186. LVOVSKY, *supra* note 9, at 11.

187. *Id.* at 11–12, 25.

188. *Id.* at 45.

189. *Id*.

190. *Id.*

191. *Id.*

192. *Id.* at 107, 156–58.

193. *Id.* at 45–47.

outsourcing firms, and once they pass a written test and an interview, they start working from home in several-hour shifts almost immediately.[194] And although human moderators are based all over the world, many of those moderating U.S. content are located in the Philippines because platforms presume that decades of U.S. colonial influence have accustomed Philippine citizens to U.S. sensibilities.[195] Therefore, like anti-vice policing, which presumed most people had access to a shared sensibility about queerness, the mechanics of content moderation implicitly rely on there being a shared understanding of what people in the United States find offensive and that anyone vaguely cognizant of U.S. norms is qualified to be a content moderator. This presumption may not have been front of mind when platforms were first designing the machinery of content moderation. Nor was this likely the primary reason for outsourcing moderation. But platforms' choice of commercial content moderators from cultures they presume are similar to those whose content they are moderating implicitly suggests that moderation is based on shared understanding about sex that ostensibly comes from cultural familiarity.

Notably, Facebook relies on several "tiers" of moderators, with commercial content moderators in the lowest tier.[196] Moderators higher up in the hierarchy often have what Facebook calls "experience judging content," whatever that actually means.[197] But Facebook is a unique case; as the largest and most profitable social network, Facebook can afford to create a system with some expertise built in. Plus, most content moderation—both at Facebook and at other platforms—happens at the entry level, with moderators in the upper tiers reviewing only a small sample of cases.[198] Therefore, for the most part, human content moderation remains the work of nonexperts.

Algorithmic moderation also relies on lay expertise. Technical expertise in building algorithms should not be confused with expertise in the substance of what those algorithms are moderating. Algorithms are models that rely on heuristics, proxies, and independent variables; they are trained on material that has been associated with previously moderated content.[199] Therefore, algorithmic moderation presumes that there are some characteristics common to sexual content, that those characteristics are evident already, and that they can be programmed into machines by

194. Chen, *Inside*, *supra* note 59; *see also* ROBERTS, SCREEN, *supra* note 7; Roberts, *Detritus*, *supra* note 7; Roberts, *Dirty Work*, *supra* note 7.

195. Chen, *Laborers*, *supra* note 73.

196. Klonick, *Governors*, *supra* note 7, at 1639–41.

197. *Id.* at 1640.

198. *Id.* at 1641.

199. Gorwa et al., *supra* note 55, at 3–5.

engineers with no firsthand experience with the underlying content. As Frank Pasquale has argued, the notion that any engineer can neatly code rules and their attendant human judgments into a machine loses the "qualitative evaluation and . . . humble willingness to recalibrate and risk-adjust quantitative data" that come with human experts.[200] Algorithmic moderation only covers code-able parts of rules against sexual content. Therefore, algorithmic moderation embodies an epistemic error: It assumes that sexual content is reducible to factors that AI can identify. Put another way: "Sex: There's an App for That." Algorithmic moderation requires no contextual or nuanced understanding of sex, let alone an appreciation for queer culture.

3. Disproportionate Antiqueer Enforcement

Given the discretion that vague rules afford algorithmic and human rule enforcers and the underlying assumptions that queerness and sexual content are readily identifiable, it should come as no surprise that both anti-vice laws and sexual content moderation rules have been disproportionately and arbitrarily applied against queer and nonnormative sexual content. Indeed, the similarities in the mechanics of enforcement run deep, even beyond the lived experiences of victims. Both anti-vice and content moderation enforcement rely on vigilantes to support surveillance. And both save the lion's share of their energy for the most marginalized within the queer community.[201] This section describes each parallel in turn.

a. Asymmetrical Enforcement: Data and Experiences

Historians have ably demonstrated the stark disproportionate application of anti-vice laws against queer people. George Chauncey found a pattern of evidence showing that New York police would arrest groups of men congregating together on certain streets or in certain spots in Central Park but ignore open and obvious indecent exposures among men who had children in tow.[202] The police raided the queer-friendly Riis

---

200. Pasquale, *Professional*, *supra* note 65, at 74; FRANK PASQUALE, NEW LAWS OF ROBOTICS 1–11 (2020); *see also* Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1 (2019).

201. Those who sit at the intersection of "matrices of domination" often face compounding subordination. *See* PATRICIA HILL COLLINS, *Black Feminist Thought in the Matrix of Domination*, *in* BLACK FEMINIST THOUGHT: KNOWLEDGE, CONSCIOUSNESS, AND THE POLITICS OF EMPOWERMENT (1990); Kimberlé Crenshaw, *Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color*, 43 STAN. L. REV. 1241 (1991).

202. CHAUNCEY, *supra* note 97, at 183.

Beach but not the heterosexual beach immediately adjacent.[203] They applied laws against solicitation to receptive partners but not to insertive partners when the latter were masculine presenting.[204] States went after queer bookstores with gay erotica but not bookstores with heterosexual pornography.[205] And documents from the New York SLA reveal that the agency saw queer and queer-friendly bars as per se "disorderly".[206] New York imposed harsher sentences on queer people in general and, specifically, on queer men convicted of loitering for cruising in a public park.[207]

Like anti-vice legislation, the rules governing sexual content moderation have been disproportionately applied to queer content.[208] Surveys of content creators show that queer people, women, persons of color, and sex workers are most likely to report that their content has been removed by major platforms.[209] Research across YouTube found that uploads featuring the words "gay" or "lesbian" or "LGBTQ"—hardly sexual or offensive—were systematically restricted by YouTube's AI.[210] Consequently, more than one-third of all videos "with queer content in the titles" were automatically demonetized, disproportionately stripping away the livelihoods of queer creators and artists.[211]

But even stories of YouTube restricting queer videos or Instagram

---

203. *See id.* at 184.

204. *Id.* at 186.

205. Eskridge, *Challenging*, *supra* note 94, at 893.

206. CHAUNCEY, *supra* note 97, at 337.

207. RYAN, *supra* note 168, at 72; MOGUL ET AL., *supra* note 140, at 78. This pattern continues. *See* Carpenter & Marshall, *supra* note 166, at 13–15 (highlighting enhanced profiling of transgender individuals); FRANK H. GALVAN & MOHSEN BAZARGAN, WILLIAMS INST., INTERACTIONS OF LATINA TRANSGENDER WOMEN WITH LAW ENFORCEMENT 7 (2012), https://williamsinstitute.law.ucla.edu/wp-content/uploads/Latina-Trans-Women-Law-Enforcement-Apr-2012.pdf [https://perma.cc/3KG3-TM9D]; Jordan Blair Woods, *LGBT Identity and Crime*, 105 CALIF. L. REV. 667 (2017); Naomi G. Goldberg, Christy Mallory, Amira Hasenbush, Lara Stemple & Ilan H. Meyer, *Police and the Criminalization of LGBT People*, *in* THE CAMBRIDGE HANDBOOK OF POLICING IN THE UNITED STATES (Tamara Rice Lave & Eric J. Miller eds., 2019).

208. Much of this evidence is ethnographic because robust quantitative and statistical analysis is impossible given platforms' refusal to share data with researchers. See *infra* section IV.B for a response to this problem.

209. *Exclusive: An Investigation into Algorithmic Bias in Content Policing on Instagram*, SALTY (Oct. 2019), https://saltyworld.net/algorithmicbiasreport-2/ [https://perma.cc/837N-XU9D].

210. BEURLING OCELOT AI, DEMONETIZATION REPORT, https://docs.google.com/document/d/18B-X77K72PUCNIV3tGonzeNKNkegFLWuLxQ_evhF3AY/edit (last visited June 20, 2021).

211. Aja Romano, *A Group of YouTubers Is Trying to Prove the Site Systematically Demonetizes Queer Content*, VOX (Oct. 10, 2019, 9:40 AM), https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report [https://perma.cc/BZV2-R2AB].

taking down queer art miss the stark inequities. In its effort to restrict "mature" or "sensitive" content, YouTube has slowed the spread and hidden much of some queer content creators' work for violating its nudity and sexual content policies; meanwhile, it has left highly sexualized content from cisgender heterosexual women untouched.[212] Queer brands, particularly those working in queer sexual health, cannot run ads on Instagram, but Necessaire and Maude, sex care brands catering to cisgender heterosexual men and women, are permitted to post revealing images and link directly to their stores.[213] When the queer band Unsung Lilly tried posting an ad to Facebook that pictured two women's foreheads touching, Facebook banned it, calling the photo "sexually explicit."[214] The band then uploaded the same ad with two different photos, one with a platonic photo of themselves and the other of a man and a woman touching foreheads. Both were approved by Facebook.[215]

Facebook allowed an ad for erectile dysfunction medication featuring a seminude man, but it blocked an ad with an image of a same-sex couple depicting the backs of their heads touching.[216] Instagram allowed the Museum of Pizza to post an image of a muscular man in briefs biting a slice of pizza, but banned an ad featuring a fully clothed man because it linked to a sex toy.[217] These are just a few examples; there are too many to list here. The result is that nonnormative and queer content is largely devoid of sexuality and sexual desire, but content that follows heterosexual norms can be sensual, racy, and explicit.[218] This is not to say that *no* queer content appears on platforms. But a disproportionate amount of queer content is restricted or blocked entirely.

b. Community Enforcement

Most histories of anti-vice policing naturally focus on the police, but government actors were aided by civilian morality groups. Some of them

---

212. Hunt, *supra* note 95.

213. Mary Emily O'Hara, *Queer and Feminist Brands Say They Are Being Blocked from Running Ads on Instagram and Facebook*, MTV (July 19, 2019, 8:00 AM), http://www.mtv.com/news/3131929/queer-and-feminist-brands-say-they-are-being-blocked-from-running-ads-on-instagram-and-facebook/ [https://perma.cc/T4ZZ-3AJY].

214. Sera Golding-Young, *Facebook Blocked My Ad, Mislabeled It "Sexually Explicit"*, ACLU N. CAL. BLOG (Sept. 23, 2020), https://www.aclunc.org/blog/facebook-blocked-my-ad-mislabeled-it-sexually-explicit [https://perma.cc/PGL9-96MB].

215. *Id.*

216. APPROVED, NOT APPROVED, https://approvednotapproved.com/ [https://perma.cc/9S6Q-B8MA] (providing comparison examples of advertising and asking users to choose which they think was allowed or not allowed on the platform).

217. *Id.*

218. Southerton et al., *supra* note 126, at 921.

were vigilantes, eager to intimidate and silence queer people through violence and assault.[219] Far more were willing to bring queer behavior to the police's attention, inform on queer-friendly bars to liquor regulators, and, in some cases, supplement police forces with their own people.[220] Sexual content moderation also depends on community members informing platforms about content they find offensive. This system may not have been designed to discriminate against queer people, but it necessarily subjects queer people to the normative judgments of others.

In the anti-vice context, community enforcement started with civilian morality clubs that engaged in public advocacy and direct action against degeneracy, solicitation, and homosexuality.[221] But much of police's help came from individuals and businesses. In the 1950s and 1960s, vice squads often responded to complaints from people and business owners.[222] Bar owners and bartenders also collaborated with liquor authorities directly.[223] Because society assumed that a queer person could be identified on sight based on any layperson's judgment, bar owners who believed they recognized queer crowds often reported them to the police.[224] The system incentivized this kind of surveillance; if the bar did not inform the city of queer clientele, the police could shut the business down for being "disorderly."[225]

This collaboration has a direct parallel in sexual content moderation, which encourages users to "flag" content for violating rules against sexual activity.[226] Flags are enforcement mechanisms; they provide "a practical mechanism for addressing the daunting task of regulating" the vast scale of platform content.[227] In other words, like anti-vice squads, which were also understaffed relative to the space they had to police, social media companies rely on members of their communities to help police their massive amounts of content and enforce their rules. Flagging also offers platform content moderation a legitimacy dividend. When social media platforms decide to remove content, they can "claim to be curating on behalf of their user community and its expressed wishes" through the flag.[228]

---

219. CHAUNCEY, *supra* note 97, at 180–81.

220. Eskridge, *Privacy*, *supra* note 160, at 717.

221. CHAUNCEY, *supra* note 97, at 138, 146, 215–16, 336; RYAN, *supra* note 168, at 89.

222. LVOVSKY, *supra* note 9, at 104.

223. *Id*. at 34–35.

224. *Id*.

225. *Id*.

226. Crawford & Gillespie, *supra* note 4, at 412.

227. *Id*.

228. *Id.*

At the same time, deputizing other users into enforcing sexual content moderation rules necessarily exposes queer content to evaluation under the terms set by dominant, normative values.[229] Michelle White found that Craigslist users leveraged the site's flagging mechanism to attack personal ads that included queer or nonnormative sexual expression.[230] Conservative groups have organized to flag pro-queer groups on Facebook en masse.[231] Flagging can turn into "user-generated warfare," where systematic flagging of queer posts for violating rules against sexual activity results in the disproportionate silencing of nonnormative queer voices.[232]

That said, there are important differences between community vice enforcement and flagging. In the anti-vice context, morality clubs operated locally and community enforcers lived and worked near queer-friendly bars.[233] They had to see a queer person, hear a disturbance, and call a local precinct. On the other hand, anyone, from anywhere, and for any reason can weaponize a flag. Viral anti-queer campaigns are easier, faster, and cheaper to organize. That vice squads never had this kind of auxiliary force even in their heyday further emphasizes sexual content moderation's capacity to put queer content at risk.

c. Intersectional Injustice

Even within the queer community, some voices are silenced more frequently than others. Indeed, both anti-vice policing and sexual content moderation share a tendency to attack the most marginalized, the presence of whom, by virtue of their intersectional identities, is automatically a threat to traditional norms.

During the height of anti-vice policing in the middle of the last century, police primarily targeted queer working-class men, queer immigrants of color, drag queens, and those who could not hide their sexual expression.[234] The historian Anne Gray Fischer has shown how the discourse of homosexuality as a mental illness was used to rationalize and

---

229. Southerton et al., *supra* note 126, at 922.

230. MICHELE WHITE, BUY IT NOW: LESSONS FROM EBAY 204 (2012) ("[T]he site's structure and flagging processes enable members to remove what they identify as not fitting or conforming.").

231. Michelangelo Signorile, *Truth4Time, Secret Religious Right Facebook Group, Included NOM Co-Founder, Fox News Pundit and More*, HUFFPOST (Dec. 6, 2017, 11:49 AM), http://www.huffingtonpost.com/2012/04/24/truth4time-secret-religious-right-facebook-group-_n_1449027.html [https://perma.cc/V57L-X9NC].

232. Brittany Fiore-Silfvast, *User-Generated Warfare: A Case of Converging Wartime Information Networks and Coproductive Regulation on YouTube*, 6 INT'L J. COMMUC'N 1965, 1973 (2012).

233. LVOVSKY, *supra* note 9, at 104.

234. CHAUNCEY, *supra* note 97, at 331–48.

mitigate responsibility for wealthy white men's same-sex behavior, but rarely, if ever, used to lessen the sentence of queer people of color and the poor.[235] Emily Hobson's masterful work on the policing of queer people in Los Angeles also has shown how police chose to arrest queer people of color more often, for more severe crimes that carried higher sentences, and with little interest in leniency.[236] Courts were more willing to show leniency to white, middle-class queer defendants, but not to working class gay men, queer immigrants, and persons of color.[237]

This biased enforcement was only partly a feature of the presumption that a queer person could be identified on sight. Granted, it was easier for police to spot the more feminine-presenting men, the more masculine-presenting women, the drag queens, those who subverted gender norms, and those bold enough to express their authentic selves.[238] Importantly, though, the discrimination also reflected the overarching goal of anti-vice policing. Chauncey notes that the state's goal was not to eradicate queer life, but rather force underground nonnormative expressions of sexuality that challenged middle-class sexual norms.[239] Wealthy, white, cisgender, masculine-presenting men could fly under the radar.[240] When they were caught up in an anti-vice sweep, they often had the family connections and wealth to keep their names out of the papers and off the arrest rolls.[241] In addition, conservative lawmakers linked their assault on queer people with their "massive resistance" to racial integration after the Supreme Court's decision in *Brown v. Board of Education*.[242] The result was a

235. Anne Gray Fischer, *"Land of the White Hunter": Legal Liberalism and the Racial Politics of Morals Enforcement in Midcentury Los Angeles*, 105 J. AM. HIST. 868 (2019).

236. Emily Hobson, *Policing Gay LA: Mapping Racial Divides in the Homophile Era, 1950–1967*, *in* THE RISING TIDE OF COLOR 188 (Moon-Ho Jung ed. 2014).

237. LVOVSKY, *supra* note 9, at 6–7, 107–09; *see also* Eskridge, *Challenging*, *supra* note 94, at 832 ("[T]he same magistrate who had reinstated the charges against [two Latino and African American men] had dismissed charges against two white men for precisely the same conduct (oral sex) with the same blond man."). This unequal policing continues. *See* AMNESTY INTERNATIONAL USA, STONEWALLED: POLICE ABUSE AND MISCONDUCT AGAINST LESBIAN, GAY, BISEXUAL AND TRANSGENDER PEOPLE IN THE U.S. 27–28, 34–36 (2005).

238. LVOVSKY, *supra* note 9, at 40 ("[E]ntertainments like the pansy cabarets . . . marked that body as *legible*, marked by a series of conspicuous visual codes.") (emphasis in original); *id*. at 41–52 (showing how knowledge of those stereotypes was used in vice policing).

239. CHAUNCEY, *supra* note 97, at 180, 333, 357.

240. LVOVSKY, *supra* note 9, at 107 ("'[N]ormal' homosexuals, indistinguishable from other men, rarely caught the vice squad's eye; it was 'the ones who dress or act aggressively or outrageously that attract[ed] our attention.'").

241. A classic example of this is the effort by Samuel Warren to keep his queer younger brother's name out of the papers and away from the police. Charles E. Colman, *About Ned*, 129 HARV. L. REV. F. 128 (2016).

242. 347 U.S. 483 (1954); NUMAN V. BARTLEY, THE RISE OF MASSIVE RESISTANCE: RACE AND POLITICS IN THE SOUTH DURING THE 1950'S 67–81, 190–236, 341–43 (1969).

disproportionate attack on alleged "sexual deviancy" at historically Black colleges and universities and Black men in general.[243]

Although there has been no systematic intersectional analysis of content moderation of queer content, there is anecdotal evidence that a similar asymmetry plagues content moderation. Content moderation algorithms find innocuous sentences that used the phrases "Black man," "Black woman," and "homosexual man" as more "toxic," and thus more likely to be automatically blocked or moderated, than the same sentences using the phrase "French man" or "German man."[244] Algorithms designed to detect hate speech have a difficult time distinguishing between homophobic and transphobic slurs and use of similar terms by queer people as a form of empowerment, putting at risk the speech of those who cannot or refuse to hide their authentic selves.[245] Users have reported that platforms hide content related to gender transition and gender-affirming treatment.[246] Instagram has removed topless images of plus-sized Black women with their arms covering their breasts ostensibly because the images violated the platform's rules against sexual content, but it does not subject the same censorship to thin white women in similar poses.[247] And in all cases of content moderation, those with money, privilege, and power are more likely to have access to legal counsel and the time and capacity to navigate platforms' appeal processes.[248] And yet, even when they do have access, social media follow another example set by anti-vice policing of queer behavior: denial of due process.

4. *Without Due Process*

Contrary to the hopes of some scholars, platform mechanisms that allow users to "appeal" unfavorable algorithmic moderation decisions do

243. MOGUL ET AL., *supra* note 140, at 38–39, 49–50.

244. Nicolas Kayser-Bril, *Automated Moderation Tool from Google Rates People of Color and Gays as "Toxic"*, ALGORITHM WATCH, https://algorithmwatch.org/en/automated-moderation-perspective-bias/ [https://perma.cc/DLC6-THCX].

245. Ben Bours, *Facebook's Hate Speech Policies Censor Marginalized Users*, WIRED (Aug. 14, 2017, 7:00 AM), https://www.wired.com/story/facebooks-hate-speech-policies-censor-marginalized-users/ [https://perma.cc/3ZRQ-H6GE].

246. Erlick, *supra* note 3.

247. Nosheen Iqbal, *Instagram 'Censorship' of Black Model's Photo Reignites Claims of Race Bias*, GUARDIAN (Aug. 9, 2020, 2:13 AM), https://www.theguardian.com/technology/2020/aug/09/instagrams-censorship-of-black-models-photo-shoot-reignites-claims-of-race-bias-nyome-nicholas-williams [https://perma.cc/4TER-ND5G].

248. *See* Marc Galanter, *Why the "Haves" Come Out Ahead: Speculations on the Limits of Legal Change*, 9 L. & SOC'Y REV. 95 (1974).

not come with added transparency.[249] This is unsurprising. Anti-vice officers routinely denied due process to their queer arrestees as well.

Arrest records and police reports from the 1930s through the 1960s demonstrate that the state neither needed a reason to raid queer-friendly bars nor informed those captured in anti-vice sweeps of the charges against them. New York's State Liquor Authority (SLA) conducted episodic raids without any justification or reason; queer men were arrested based on single, uncorroborated reports of loitering.[250] During the Lavender Scare, the Civil Service Commission fired queer government employees for "immoral conduct" without any specific evidence, rationale, or warning.[251] Police often failed to apprise queer defendants—particularly men of color—of their rights to silence and counsel.[252] This continues today, where transgender women are arrested without reason for congregating or walking on the street at night.[253]

Denial of due process is just one factor that delegitimized anti-vice policing.[254] But it was an integral part of the assemblage of forces that exerted power over queer identity and expression. Refusing to inform an arrestee of the reason for their arrest, harassing law-abiding citizens on the street, and detaining people for "walking while queer" both reflect and reify the extent to which police denied the humanity and dignity of their targets. And if, as legal historians suggest, the goal of anti-vice policing was to push queer sexuality back into the closet, the threat of arbitrary arrest without traditional due process protections likely had a powerful chilling effect on queer people.[255]

Following this tradition, social media platforms do not provide their users with explanations for adverse content moderation. Algorithmic decisions are mostly opaque, and human reviews operate in a black box.[256] When individuals disagree with a content moderation decision, some social media companies provide an "appeal" option, which constitutes the

249. *See* Bloch-Wehba, *supra* note 7, at 91.

250. CHAUNCEY, *supra* note 97, at 170–71, 341–42.

251. *See* Scott v. Macy, 349 F.2d 182, 185 (D.C. Cir. 1965).

252. Eskridge, *Challenging*, *supra* note 94, at 832.

253. Carpenter & Marshall, *supra* note 166, at 6; MOGUL ET AL., *supra* note 140.

254. *See* TOM R. TYLER, WHY PEOPLE OBEY THE LAW (2d ed. 2006); Tom R. Tyler, *Governing Amid Diversity: The Effect of Fair Decisionmaking Procedures on the Legitimacy of Government*, 28 L. & SOC'Y REV. 809 (1994).

255. *E.g.*, Steffel v. Thompson, 415 U.S. 452, 458–60 (1974) (suggesting that the chilling effect when a pamphleteer is threatened with arrest for continued distribution of handbills is sufficient for Article III standing).

256. Roth, *supra* note 88, at 415, 418, 423.

extent of the process involved in challenging platform decisions.[257] Users have described this process as "speaking into a void."[258] Platforms rarely provide space for either explanation or context. Users usually only receive a final decision.[259] Despite platforms' eagerness for the legitimacy dividend that comes with being treated as a quasi-government, it is difficult to characterize this as an "appeal"; after all, moderators must make decisions in seconds, without the benefit of deliberation, context, and analysis.[260]

C. Sexual Content Moderation's Path Dependencies

This Article has argued that similar assemblages of sociolegal forces, ideas, rules, technologies, people, and machinery characterize both sexual content moderation and anti-vice policing that targeted the queer community from the 1930s through the 1960s. These assemblages amplify power structures, whether it was municipal leaders and police chiefs eager to reassert government control over urban nightlife or technology companies looking to establish highly profitably control of their platforms without public accountability. The effect of both assemblages is a disproportionate crackdown on queer people and expressions of sexual nonnormativity.

To be clear, this analogy does not imply that anti-vice policing and sexual content moderation were developed with the specific and malicious intent of discriminating against queer people. That is not how assemblages work. An assemblage is an aggregation of social forces that "'work' together as a functional entity."[261] Kevin Haggerty and Richard Ericson identified a "surveillant assemblage" of different technologies, institutions, people, and knowledge—themselves also comprising constituent parts—that come together to subject people to the kind of total surveillance that any one set of them could not.[262] Electronic monitoring is a good example. The practice does not just involve ankle bracelets; it also consists of phone calls, programmed contact, reporting tracking stations, databases, satellites, electricity, movement and voice, and so

257. Kristen Vaccaro, Christian Sandvig & Karrie Karahalios, *"At the End of the Day Facebook Does What It Wants": How Users Experience Contesting Algorithmic Content Moderation*, 4 PROC. ACM HUM.-COMPUT. INTERACTION 167:1, 167:4 (2020).

258. Sarah Myers West, *Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms*, 20 NEW MEDIA & SOC'Y 4366, 4378 (2018) (quoting user report).

259. Southerton et al., *supra* note 126, at 923.

260. GILLESPIE, *supra* note 7, at 7; Roberts, *Detritus*, *supra* note 7.

261. Haggerty & Ericson, *supra* note 9, at 608.

262. *Id.* at 608–09.

on.[263] Individually, these forces did not set out to create hegemonic systems of surveillance; as they "transcend[] institutional boundaries, systems intended to serve one purpose find other uses."[264] More than just surveillance creep, where a single technology is repurposed to engage in additional monitoring,[265] the surveillant assemblage consists of different pieces that exert power in a way greater or different than the sum of its parts.

That is what's happening in the content moderation space. Like the assemblage of laws, ideas, rules, technologies, expertise, people, and processes that constituted anti-vice policies, the sexual content moderation assemblage results in the disproportionate censorship of queer expression because of the path dependencies of its parts working as they were designed. Sexual content moderation was built to keep pornography away from children, to keep advertising profitable, and to ensure institutional control over platforms.[266] Because *sexual* nonnormativity is one defining feature of queerness, queer content is disproportionately swept up in that machinery. Facebook and Instagram algorithmically restrict and slow the dissemination of content with hashtags like #gays, #iamgay, #lesbiansofinstagram, and #lesbian because the hashtags are often spammed with pornography.[267] Users erase queer language or replace it with emojis or symbols because using certain sexually tinged words triggers algorithmic moderation.[268] Researchers at the Australian Strategic Policy Institute found that TikTok was shadow banning (or, making it difficult to find) queer hashtags—#gay, #transgender, and #Iamagay/lesbian—in Russia, Bosnia, and Jordan, and in at least eight languages to please politicians and advertisers.[269]

Platforms wrote vague rules to provide flexibility and discretion. They outsourced most of the work to algorithms, as well as commercial content moderators in the Global South and other users when the sheer scale of

263. *Id.* at 610.

264. *Id.* at 616.

265. Matthew Tokson & Ari Ezra Waldman, *Social Norms in Fourth Amendment Law*, 120 MICH. L. REV. 265, 301–02 (2021).

266. *See supra* section II.A.2.

267. Erlick, *supra* note 3.

268. Roth, *supra* note 88, at 427–28.

269. Fergus Ryan, Audrey Fritz & Daria Impiombato, *TikTok and WeChat: Curating and Controlling Global Information Flows*, AUSTL. STRATEGIC POL'Y INST. (Sept. 8, 2020), https://www.aspi.org.au/report/tiktok-wechat (last visited Nov. 18, 2022); Jane Li, *TikTok Is Suppressing LGBT Content in Eastern Europe and the Middle East*, QUARTZ (Sept. 8, 2020), https://qz.com/1900530/tiktok-shadow-bans-lgbt-hashtags-in-russian-and-arabic [https://perma.cc/NMA9-5UVZ].

content became too massive.[270] However, even though outsourcing and algorithmic moderation arose to address the problem of scale, their operation disproportionately suppresses queer content because it subjects queer expression to lay decisions made according to normative traditions and mainstream perceptions of "common sense." That is, coders likely did not set out to design anti-queer moderation algorithms. Nor is it likely that most commercial content moderators are maliciously censoring queer content. But the system relies on decisionmakers who not only reflect normative traditions, but also come with no particular expertise in content, sexuality, and queer culture. Like anti-vice policing, the result is that in the course of achieving certain economic and institutional goals, sexual content moderation ends up disproportionately impacting queer expression.

This has profound implications for how scholars understand content moderation. The literature tells us that content moderation reflects free speech norms, systemic balancing of interests, and even the processes of administrative law.[271] This Article's analogy suggests that these arguments are, at best, ex post justifications for a phenomenon that actually replicates systems of control that have long been used to police public expressions of queer sexuality. The architects of content moderation at the Internet's largest platforms may not want to admit it. They may not even appreciate their complicity. But their work has the unmistakable fingerprints of oppressive and subordinating regimes.

D. Limitations

That said, this analogy has its limits. Some of those limits are obvious. Unlike anti-vice policing, content moderation arose amidst conflicts between private platforms and the state over liability and the reach of law. As Julie Cohen has argued, those conflicts contributed to the development of performative practices of moderation "designed to express a generic commitment to accountability without . . . enabl[ing] meaningful scrutiny of the underlying processes."[272] In addition, punishments meted out by the state are qualitatively different from throttling (slowing down content), shadow banning, and takedowns. The latter can destroy people's livelihoods, silence unpopular opinions, and negatively affect marginalized populations.[273] But some queer men convicted of sodomy

270. *See supra* sections II.B.1–2.

271. Klonick, *Governors*, *supra* note 7; Douek, *Governing*, *supra* note 6; Douek, *Systems*, *supra* note 13.

272. COHEN, *supra* note 15, at 250.

273. Elkin-Koren et al., *supra* note 32, at 6, 13–15.

spent decades or more in jail.[274] Others were forced to register as sex offenders; some had their names published in local papers.[275]

It is also difficult to compare the lived experiences of queer people in these two different eras. Homophobia and transphobia remain rampant, dangerous, and violent. But those with privilege—white, cisgender, masculine-presenting men and feminine-presenting women—do not always walk in public with the same level of fear that they and their elders did even two decades ago, let alone nine. Plus, the lived experiences of queer people and other members of marginalized groups suggest that enforcement by algorithm or commercial content moderators is not the same as enforcement by the police. Scholars have argued that law enforcement is institutionally committed to policing sexual and gender "deviance."[276] Its racial biases put the queer community's most marginalized—queer Black women, Black transgender women, and nonbinary persons of color—at particular risk.[277] The problem is systemic.[278] Therefore, for many queer people, points of contact with police are unique moments of danger and violence that are not recreated in the context of sexual content moderation.

Despite these and the limitations discussed throughout this Part, understanding sexual content moderation through the anti-vice lens is a useful way to uncover the politics, values, and power of platforms. The parallel assemblages of power have similar effects. In their quest to exert control and achieve their own economic, moralistic, and institutional goals, both anti-vice policing and sexual content moderation disproportionately harm queer people and silence their voices.[279] Even more, this Article's historical parallel teaches us something new about content moderation elided in the current literature. That is, like anti-vice policing, the result of sexual content moderation is the creation and reification of social media as straight spaces. The next Part demonstrates how.

III. SOCIAL MEDIA AS A "STRAIGHT SPACE"

Having argued that sexual content moderation is best understood as an

---

274. LVOVSKY, *supra* note 9, at 196.

275. Eskridge, *Challenging*, *supra* note 94, at 819.

276. MOGUL ET AL., *supra* note 140, at xiii, 47.

277. *See* sources cited *supra* note 201.

278. Jerome Hunt & Aisha C. Moodie-Mills, *The Unfair Criminalization of Gay and Transgender Youth*, CTR. FOR AM. PROGRESS (June 29, 2012), https://www.americanprogress.org/issues/lgbt/reports/2012/06/29/11730/the-unfaircriminalization-of-gay-and-transgender-youth/ [https://perma.cc/T7TP-GF5B].

279. Southerton et al., *supra* note 126, at 921.

assemblage of social forces that resembles anti-vice policing, this Part takes the next step. If the path-dependent result of the sexual content moderation assemblage is the disproportionate censorship of queer and nonnormative expression, then sexual content moderation contributes to the creation and maintenance of social media as a "straight space." The concept of a straight space is based on Elijah Anderson's concept of "white space"—namely, those "neighborhoods, restaurants, schools, universities, workplaces, churches and other associations, courthouses, and cemeteries . . . that reinforce[] a normative sensibility in settings in which black people are typically absent, not expected, or marginalized when present."[280] White spaces can make those who identify as Black or African-American "feel uneasy and consider it to be informally 'off limits.'" On the other hand, white people see white settings as "unremarkable, or as normal, taken-for-granted reflections of civil society."[281]

As I am adapting the term, straight spaces are built on heterosexual norms, constructed to be unwelcome to queer and nonnormative expression, and designed to reify the heteronormative and cisgender supremacy of our institutions. Like white spaces, straight spaces are not entirely off limits to those who identify in other ways. Nor are white and straight spaces static.[282] White and straight spaces can be spaces of resistance for both the historically oppressed, and even spaces with many queer people can still be straight spaces if they are built on foundations of heteronormativity.[283] But like white spaces, straight spaces have the effect of reifying dominant paradigms. And, just like Black people "are required to navigate the white space as a condition of their existence," queer people are forced to live in straight spaces.[284]

This Part relies on similarities and differences between sexual content moderation and anti-vice policing to show how the practices, technologies, and background law of sexual content moderation constructs social media as a straight space. First, I describe how both anti-vice policing and sexual content moderation deprive individual queer people of "sexual citizenship" and contribute to discourses of discrimination. Second, I show how the technologies and capacities of

280. Anderson, *supra* note 14, at 10.

281. *Id.*

282. *Id.* at 10–11.

283. Although not using the phrase "straight space," Yoel Roth has argued that supposedly queer spaces like Grindr, the geosocial dating and hook up app used mostly by gay and bisexual men, still marginalize queerness and keep "[n]onnormative practices—fetishistic, 'unsafe,' or highly visible sexualities, for instance— . . . consistently hidden from view." Roth, *supra* note 88, at 429.

284. Anderson, *supra* note 14, at 11.

both anti-vice policing and sexual content moderation have the effect of designing spaces to be unwelcoming to queer content. Finally, I describe how the background law of content moderation entrenches heteronormative values in moderated social spaces online.

A. Stigmatizing Queerness

Anti-vice laws enforced a normative sexual hierarchy, where queer people were at the bottom. As this Article's analogy shows, the same is true for social media's sexual content moderation assemblage. By trying to keep pornography away from children and sanitize the marketplace for advertisers, sexual content moderation reifies heteronormative traditions and controls what queer people and the wider population understand to be queer identity. This harms queer content creators, skews popular discourse, and justifies discrimination online.

Gayle Rubin argued that conflicts over sex all have a similar hierarchical dynamic. On one side of the sex ledger is "good," "normal," and "natural" sex: heterosexual, monogamous, procreative, noncommercial, same generational, and in private, among others characteristics.[285] On the wrong side of the hierarchy are nonnormative practices: homosexual, promiscuous, hedonistic, cross-generational, in public, commercial, and any other sexual practices not captured.[286] Anti-vice policing identified archetypes of "bad" sexual behavior, including sexual crimes against children.[287] In an effort to police such "bad" sexual behavior, vice squads eventually started arresting men who just looked queer.[288] Similarly, in an effort to police some of the "bad" content—namely, child pornography, sexual exploitation, and nonconsensual pornography—to make platforms business- and family-friendly, sexual content moderation sweeps up queer content as well. This markets queerness as taboo, shameful, and unacceptable.[289]

This is the assemblage at work. Sexual content moderation's rules may not have been intentionally designed to discriminate against queer content. Nonetheless, platforms developed them in a conservative sociolegal context, designed them to achieve corporate goals, and allowed them to give discretion to nonexpert enforcers who often have normative

285. Gayle Rubin, *Thinking Sex: Notes for a Radical Theory of the Politics of Sexuality*, *in* CULTURE, SOC'Y AND SEXUALITY 143, 152–53 (Richard Parker & Peter Aggleton eds., 2d ed. 2007).

286. *Id*. at 153–54.

287. LVOVSKY, *supra* note 9, at 5; JOHNSON, *supra* note 108, at 73.

288. *See* LVOVSKY, *supra* note 9, at 11–12.

289. MICHAEL WARNER, THE TROUBLE WITH NORMAL: SEX, POLITICS, AND THE ETHICS OF QUEER LIFE 22–26 (Harvard Univ. Press 1999).

sexual biases. As a result, these rules conflate queer sexuality with illicit sexuality.

This has two principal effects. First, for queer content creators whose work is restricted or taken down, sexual content moderation deprives them of their "sexual citizenship."[290] Sexual citizenship refers to "sexual claims of belonging" or one's ability to participate in society fully with authentic and honest expressions of sexual identity.[291] The disproportionate censorship of queer content denies queer people the opportunity to exist on social media with the same chance as others to share, advocate, and live authentically.[292]

Second, this censorship also systemically stigmatizes queerness in general. Anti-vice laws' asymmetrical enforcement against queer people branded queerness as "degenerate" or "lewd" or "disorderly," justifying denial of services in areas beyond urban nightlife.[293] As George Chauncey argued, the sexual repression typified by anti-vice policing also trivialized queer people's experiences, justifying institutional and social discrimination long after the riots at the Stonewall Inn and the systematic dismissal of queer dignity during the AIDS crisis.[294] For queer people in general, then, rules that put their experiences on the "wrong" side of the sexual hierarchy are first steps in a larger anti-queer agenda. Sexual content moderation is just the latest force perpetuating that stigmatization. Dominant platforms' tendency to only make accessible a sanitized version of queer life stigmatizes nonnormative expressions of sex, sexuality, and gender identity, giving space for discrimination on the basis of those characteristics and behaviors.

That is because social media platforms and their content moderation

---

290. Southerton et al., *supra* note 126, at 921.

291. Peter Aggleton, Sujith Kumar Prankumar, Rob Cover, Deana Leahy, Daniel Marshall & Mary Lou Rasmussen, *Introduction*, *in* YOUTH, SEXUALITY AND SEXUAL CITIZENSHIP 6 (Peter Aggleton et al. eds., 2019); *see also* DAVID EVANS, SEXUAL CITIZENSHIP: THE MATERIAL CONSTRUCTION OF SEXUALITIES (1993) (coining the phrase "sexual citizenship").

292. Sexual citizenship resembles Danielle Citron's concept of sexual privacy in that both focus on the dignitary benefits of control over sexual selves and expression. *See* Citron, *Sexual Privacy*, *supra* note 7.

293. Anti-vice laws are, of course, not the only examples of this. For instance, persistent anti-sodomy laws, even when they were rarely enforced, not only categorized all gay people as presumptive criminals, but also justified many other forms of discrimination. *See, e.g.*, Nan Hunter, *Sexual Orientation and the Paradox of Heightened Scrutiny*, 102 MICH. L. REV. 1528, 1536 (2004) (comparing adultery laws to sodomy laws in that both helped "solidify[] social and cultural norms"); Christopher Leslie, *Creating Criminals: The Injuries Inflicted by "Unenforced" Sodomy Laws*, 35 HARV. CIV. RTS.-CIV. LIBERTIES L. REV. 103, 103 (2000) (arguing that "very existence" of even unenforced sodomy laws "creates a criminal class of gay men and lesbians" justifying discrimination, harassment, and violence).

294. *See* CHAUNCEY, *supra* note 97.

rules play critical roles in the production and dissemination of knowledge. Queer oral histories were interrupted by AIDS.[295] The vast majority of us do not have access to elders who can guide us. Social media has become our guide. Far more than their physical counterparts, social media platforms are the sites of queer expression today.[296] Platforms' queer content is, therefore, how many people will come to understand the defining features of queerness, even if the content is just a highly curated slice of the real thing. If queer adolescents want to learn how to put on makeup for their first drag show, only a few have drag families to join. Those who do not go to YouTube. If someone wants to learn the best way to come out to their parents as bisexual, they join a Facebook group or watch videos on TikTok. Medical schools insufficiently prepare future physicians to understand the needs and health risks of queer people.[297] Therefore, without enough doctors trained in queer sexuality, adolescents turn to Google to learn how to safely prepare for intimacy.[298] Social media websites are today's gatekeepers of queer knowledge, and, as Michel Foucault predicted, platforms' continued repression of the queer experience establishes a skewed, heteronormative baseline understanding of what it means to be queer.[299]

A similar thing happened in the anti-vice context. As Professor

---

295. *See* NAN ALAMILLA BOYD & HORACIO ROQUE RAMIREZ, BODIES OF EVIDENCE: THE PRACTICE OF QUEER ORAL HISTORY (Nan Alamilla Boyd & Horacio Roque-Ramirez eds., Oxford Univ. Press 2012).

296. Brady Robards, Brendan Churchill, Son Vivienne, Benjamin Hanckel & Paul Byron, *Twenty Years of 'Cyberqueer': The Enduring Significance of the Internet for Young LGBTIQ+ People*, *in* YOUTH, SEXUALITY AND SEXUAL CITIZENSHIP 151, 151–67 (Peter Aggleton et al. eds., 2018); Jon Wargo, *"Every Selfie Tells a Story": LGBTQ Youth Lifestreams and New Media Narratives as Connective Identity Texts*, 19 NEW MEDIA & SOC'Y 560 (2015).

297. *E.g.*, Alyssa Wittenberg & Judith Gerber, *Recommendations for Improving Sexual Health Curricula in Medical Schools: Results from a Two-Arm Study Collecting Data from Patients and Medical Students*, 6 J. SEX MED. 362 (2009).

298. There are indeed some areas of queer life where we should want empowered medical professionals and scientific evidence to inform legal rules and personal behavior. Gender-affirming hormone therapy is a good example. Medical expertise can also help prepare young people to engage in and enjoy intimacy, especially since social media is full of misinformation about HIV and anal sex. For examples of how medical expertise can help prepare young people to engage in and enjoy intimacy, see *Anal Douching: How to Safely Clean for Anal*, BESPOKE SURGICAL, https://bespokesurgical.com/education/anal-cleansing/ [https://perma.cc/6KD7-Y23Q]; Evan Goldstein, *How Often Should You Use an Enema?*, BESPOKE SURGICAL (Oct. 5, 2020), https://bespokesurgical.com/2020/10/05/how-often-should-you-use-an-enema/ [https://perma.cc/69G7-P8WT]. But it is not obvious that medical expertise is always necessary since communities can help each other in more supportive, joyous, and democratic ways.

299. MICHEL FOUCAULT, THE ARCHEOLOGY OF KNOWLEDGE 89 (1972); Stuart Hall, *The West and the Rest: Discourse and Power*, *in* FORMATIONS OF MODERNITY 201–02 (Stuart Hall & Bram Gieben eds., 1992); *see also* Eskridge, *Challenging*, *supra* note 94, at 829 (noting that controlling knowledge and discourse can determine the course of changing attitudes and law).

Lvovsky argues, during the height of anti-vice policing, when queer expression was removed from media, banned in the mail, and pushed off the streets, police perversely became the arbiters of what it meant to be queer.[300] Law enforcement agents became the ones who could tell reporters how to "spot a gay" because that is what anti-vice laws required of them; when reporters asked, police recited stereotypes of queer life, which were then disseminated through the media.[301]

This skewed perception of queerness is still animating hate and discrimination today. Between January and May 2021, there were 250 anti-LGBTQ bills introduced in the states, many of which prohibit transgender adolescents from accessing gender-affirming hormone therapy or ban transgender kids from playing in school sports in accordance with their gender identities.[302] This follows on the heels of a raft of laws that force trans folks to use only those single-gender public bathrooms that match their sexes assigned at birth.[303] Texas now considers gender-affirming healthcare child abuse.[304] In 2020, a federal appeals court overturned a city's ban on gay conversion therapy, holding that the free speech rights of therapists predominate over the government's interest in protecting queer adolescents.[305]

When social media systematically removes queer content, it otherizes queer identity, hiding the full breadth of queer people's search for equality and making them vulnerable to this and other forms of discrimination. Inside the queer community, heated debates over nonnormative queer expression at pride festivals are likely also influenced by the narrow vision of queerness disseminated after sexual content moderation filters

---

300. LVOVSKY, *supra* note 9, at 13, 15.

301. *Id.* at 14–18.

302. Press Release, Wyatt Ronan, Human Rights Campaign, 2021 Officially Becomes Worst Year in Recent History for LGBTQ State Legislative Attacks as Unprecedented Number of States Enact Record-Shattering Number of Anti-LGBTQ Measures into Law (May 7, 2021), https://www.hrc.org/press-releases/2021-officially-becomes-worst-year-in-recent-history-for-lgbtq-state-legislative-attacks-as-unprecedented-number-of-states-enact-record-shattering-number-of-anti-lgbtq-measures-into-law [https://perma.cc/6ZRU-32JY].

303. Scott Skinner-Thompson, *Performative Privacy*, 50 U.C. DAVIS L. REV. 1673, 1709–12 (2017); Scott Skinner-Thompson, *North Carolina's Catch-22*, SLATE (May 16, 2016, 8:45 AM), http://www.slate.com/blogs/outward/2016/05/16/north_carolina_s_hb2_puts_transgender_people_in_an_impossible_catch_22.html [https://perma.cc/R4Q9-86YQ].

304. J. David Goodman & Amanda Morris, *Texas Investigates Parents over Care for Transgender Youth, Suit Says*, N.Y. TIMES (Mar. 1, 2022), https://www.nytimes.com/2022/03/01/us/texas-child-abuse-trans-youth.html [https://perma.cc/JBZ4-R9M7]. This is just the latest attack on transgender minors. In the first two months of 2022, fourteen states introduced twenty-five proposals to ban gender-affirming healthcare for minors. *Legislative Tracker: Youth Healthcare Bans*, FREEDOM FOR ALL AMS. https://freedomforallamericans.org/legislative-tracker/medical-care-bans/ [https://perma.cc/4NUW-WY8J].

305. *See* Otto v. City of Boca Raton, 981 F.3d 854 (11th Cir. 2020).

out nonnormative expression.[306] In other words, just like we cannot understand queerness and its legal status today without looking at the influence of anti-vice policing, we cannot understand ongoing debates about queer equality and liberation without exploring the influence of sexual content moderation.[307]

B. Technology in Straight Spaces

Content moderation and anti-vice policing are also case studies of technology's place in society. Both anti-vice authorities and social media platforms deployed the new technologies of their days to achieve their economic, moral, and institutional goals. And because technologies reflect the social relations in which they are designed and used, they had the further effect of erecting systemic barriers to queer participation.

Langdon Winner memorably argued that technologies have politics that embody the social relations that created them.[308] This occurs either because "the invention, design, or arrangement of a specific technical device or system becomes a way of settling an issue in the affairs of a particular community" or because the systems are "inherently political technologies," which "appear to require, or to be strongly compatible with, particular kinds of political relationships."[309] Winner argues that "to recognize the political dimensions in the shapes of technology does not require that we look for conscious conspiracies or malicious intentions."[310] There are many cases in which "the technological deck has been stacked in advance to favor certain social interests."[311] For instance, Robert Moses embedded New York's highways with discriminatory politics by placing overpasses just low enough to deny access to the

---

306. Alex Abad-Santos, *The Perpetual Discourse over LGBTQ Pride, Explained*, VOX (June 2, 2021, 9:00 AM), https://www.vox.com/the-goods/22463879/kink-at-pride-discourse-lgbtg [https://perma.cc/D9EK-PHHA]; Skylar Baker-Jordan, *BDSM and Kink Don't Belong in Pride Celebrations. This Is Why*, INDEPENDENT (May 25, 2021, 9:52 PM), https://www.independent.co.uk/voices/bdsm-kink-pride-lgbt-rights-celebrations-why-b1853859.html [https://perma.cc/BVK2-TXGV].

307. In this way, this Article builds on the liberatory work of the "data colonialism" movement, which considers Facebook's and other large platforms' oppressive capacities. *See* Adrienne LeFrance, *Facebook and the New Colonialism*, ATLANTIC (Feb. 11, 2016), https://www.theatlantic.com/technology/archive/2016/02/facebook-and-the-new-colonialism/462393/ [https://perma.cc/5YST-JJGW]; Jim Thatcher & David O'Sullivan, *Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data*, 34 ENV'T & PLAN. D: SOC'Y & SPACE 990 (2016).

308. Langdon Winner, *Do Artifacts Have Politics?*, 109 DAEDALUS 121, 123 (1980).

309. *Id.*

310. *Id.* at 125.

311. *Id.* at 125–26.

mostly poor, Black, and working class New Yorkers who had to take the bus to the beach.[312] Baron de Haussmann famously redesigned Paris with broad thoroughfares to avoid another working class uprising in which resistance fighters used household furniture to barricade Paris's narrow streets and choke off troop movements.[313] Anti-vice policing and sexual content moderation are similar. They are imbued with the politics of straight spaces even if we assume they were not designed with malicious, anti-queer intent.

Anti-vice police took surreptitious photographs of clandestine meetings of queer men, used new video surveillance technology of parks, took advantage of the relatively small size of new remote-controlled cameras to place them inside towel dispensers in public bathrooms, and then wired adjacent utility closets to listen and watch.[314] Police supplemented that technology-based surveillance with analog ones: decoys, trails, plainclothes officers, and handbooks and training manuals that helped officers identify a queer person on sight.[315] These technologies were embedded with the politics of straight spaces because they were specifically used to police queer meeting places, not straight ones. Anti-vice police had lost control of public bathrooms; they were hard to monitor with traditional means. Cruisers, as men looking for sex in public were called, recognized the relative safety that came with windowless walls and closed bathroom stalls and developed a complex system of codes to protect themselves and evade police.[316] Surveillance technologies settled the issue of whether these spaces were subject to public morality regulation by enhancing police's surveillance capacities and extending their eyes and ears into places they could never reach before. The technologies of clandestine surveillance were also deployed in a place almost exclusively used for same-sex activity. They could have been used anywhere, but historians have found that police only set up high-tech operators in public bathrooms known to be meeting places for queer men.[317] Therefore, such technologies both reflected and amplified the anti-queer politics of the anti-vice era.

Sexual content moderation may use radically more advanced technologies, but their politics are similar. Algorithmic moderation uses

---

312. *Id.* at 123–24; *see also* ROBERT CARO, THE POWER BROKER: ROBERT MOSES AND THE FALL OF NEW YORK 318, 481, 951–58 (1974).

313. DAVID PINKNEY, NAPOLEON III AND THE REBUILDING OF PARIS 35–39 (1958).

314. LVOVSKY, *supra* note 9, at 189–93.

315. *Id.* at 143–44, 186–87.

316. LAUD HUMPHREYS, TEAROOM TRADE 64–65 (1970); CHAUNCEY, *supra* note 97, at 188; LVOVSKY, *supra* note 9, at 185.

317. LVOVSKY, *supra* note 9, at 183–86.

AI-powered matching and classification to determine the propriety of user-generated content.[318] Matching involves comparing a new piece of content against an existing database of inappropriate content not allowed on the platform; classification takes new content and tries to categorize it into one or more clusters, which helps determine if the content will be allowed.[319] Both involve AI and machine learning, and both have blind spots. Databases have to be updated to remain relevant.[320] And, as noted above, these techniques often lack context to be able to classify idiomatic or nonnormative content appropriately.[321] Importantly, the machine learning techniques used to train both types of algorithms rely on historical data produced within biased institutions based on normative standards.[322] Both systems reflect the biases of their designers, most of whom represent dominant cultures—white, cisgender, heterosexual, able, and male.[323] As a result, heteronormative biases pervade algorithmic moderation, both from the data on which moderation algorithms are trained and from the designers themselves. Encoded in these technologies is the value-laden claim that queerness *is* sexual in a way that straightness is not. By bringing heteronormative biases to the design of moderation decision-making, these technologies help social media companies ensure that their platforms will remain straight spaces. As Tarleton Gillespie has noted, platforms do have politics;[324] sexual content moderation demonstrates the extent of their heteronormativity.

C. *The Role of Law*

The analogy between anti-vice policing and sexual content moderation surfaces the role that law plays in creating and maintaining social media as a straight space. Law is often complicit in denying queer access to traditionally heterosexual spaces. Zoning laws have reduced the

---

318. Gorwa et al., *supra* note 55, at 4–5.

319. *Id.* at 5.

320. *Id.*

321. *See supra* section II.B.2.

322. *See* SAFIYA NOBLE, ALGORITHMS OF OPPRESSION (2018); SARAH MYERS WEST, MEREDITH WHITTAKER & KATE CRAWFORD, AI NOW INST., DISCRIMINATING SYSTEMS: GENDER, RACE, AND POWER IN AI (Apr. 2019), https://ainowinstitute.org/discriminatingsystems.pdf [https://perma.cc/PGS5-73VK]; Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classifications*, 81 PROCEEDINGS OF MACHINE LEARNING RES. 1, 10–11 (2018); Pauline Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 874–90 (2017); Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

323. *E.g.*, Sonia Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 68–89 (2019) (detailing sources of bias in AI).

324. Tarleton Gillespie, *The Politics of Platforms*, 12 NEW MEDIA & SOC'Y 347, 349 (2010).

availability of businesses catering to queer people by redesigning neighborhoods to be "family friendly."[325] Federal courts have restricted queer access to public spaces to protect others' normative expression.[326] Universities' refusal to recognize queer affinity clubs was, at least in part, about denying access to physical spaces open to heterosexuals.[327] And school administrators who refused to integrate proms or parents who objected to inclusive sex education classes sought to deny queerness access to the physical space of the classroom.[328] This focus on space is unsurprising. Control over public spaces is control over public discourse; silencing queer expression keeps sexual norms in place.

Similarly, law was everywhere in the anti-vice context, embedding stereotypes, regulating queer behavior, and granting substantial discretion to the police to remove queer sexual behavior from public spaces.[329] But as Lvovsky insightfully demonstrates, some courts also pushed back; the use of the courts by the resistance ultimately helped eliminate some of the anti-vice era's vague liquor laws and undermined police's vast authority.[330] As we have discussed, other laws also created and enhanced social media's power to overpolice queer content. But courtrooms' doors are closed to those trying to resist the censorious efforts of private technology platforms.

Bill Eskridge has exhaustively profiled the ways queer legal advocates used due process, privacy, and equal protection arguments to challenge anti-vice laws and their disproportionate application to queer people.[331]

---

325. *See* SAMUEL DELANY, TIMES SQUARE RED, TIMES SQUARE BLUE xxi, xxii (1999); Lauren Berlant & Michel Warner, *Sex in Public*, 24 CRITICAL INQUIRY 547, 551 (1998); Marc Elovitz & P.J. Edwards, *The D.O.H. Papers: Regulating Public Sex in New York City*, *in* POLICING PUBLIC SEX 295–316 (Ephen Colter et al. eds., 1996). This is similar to law determining the location and construction of highways, which "symbolically and physically" encouraged racial segregation and disempowerment. Deborah Archer, *"White Men's Roads Through Black Men's Homes": Advancing Racial Equity Through Highway Reconstruction*, 73 VAND. L. REV. 1259, 1264 (2020) (quoting PAUL FOTSCH, WATCHING THE TRAFFIC GO BY 4 (2007)).

326. *E.g.*, Hurley v. Irish-Am. Gay, Lesbian, and Bisexual Group of Boston, 515 U.S. 557, 559 (1995) (affirming the right of organizers of the St. Patrick's Day Parade to prohibit openly gay marchers).

327. *E.g.*, Gay Student Serv's v. Tex. A&M Univ., 737 F.2d 1317, 1319 (5th Cir. 1984) (requiring public university to recognize gay-straight alliance); Gay Rts. Coal. of Georgetown Univ. L. Ctr. v. Georgetown Univ., 536 A.2d 1, 5 (D.C. Cir. 1987) (allowing university to deny recognition of queer group, but not "tangible" benefits, including space, on the basis of sexual orientation).

328. *E.g.*, Fricke v. Lynch, 491 F. Supp. 381, 387–88 (D.R.I. 1980) (rejecting school district's safety rationales for prohibiting a student from bringing a same-sex date to prom); Curtis v. Sch. Comm. of Falmouth, 652 N.E.2d 580, 583 (Mass. 1995) (denying parental objection to educating students about condom use and safe sex).

329. ESKRIDGE, GAYLAW, *supra* note 16, at 17–97; LVOVSKY, *supra* note 9, at 117.

330. LVOVSKY, *supra* note 9, at 48–49, 134–37, 196–99, 257–58.

331. Eskridge, *Challenging*, *supra* note 94, at 819.

Lawyers invoked the right to privacy to successfully challenge police searches of closed toilet stalls and other private spaces.[332] To combat anti-vice laws, they applied the Supreme Court's decision in *Papachristou v. Florida*,[333] a case where the Supreme Court invalidated an ordinance that made it a crime for people to be "vagabonds" or "lewd, wanton, and lascivious persons"; the Court reasoned that the ordinance was unintelligible and disproportionately applied to "nonconformists."[334] The petitioners in *Papachristou* were two Black men and two white women who were arrested under Florida's "vagabond" law simply for riding together in a car.[335] The same vagueness that doomed the law in that case was sometimes successfully used to put an end to vague anti-vice laws' disproportionate application against queer men.

Lvovsky describes how lawyers representing queer victims of anti-vice laws also challenged vice squad's lack of evidence and misconduct. For instance, in *Scott v. Macy*,[336] the ACLU successfully challenged the Civil Service Commission's summary firing of a queer employee for his disorderly conduct arrests. The court held that if the government wanted to fire someone for "immoral conduct," it had to specify what conduct was immoral and how that conduct affected "occupational competence or fitness."[337] In *Stoumen v. Reilly*,[338] the California Supreme Court held that the state should never have shut down the bohemian and queer-friendly Black Cat in San Francisco simply for allowing "persons of known homosexual tendencies" to gather.[339] The state needed to prove actual "illegal or immoral acts" to shut down a bar for maintaining a "disorderly" house.[340] And in *Kelly v. United States*,[341] the D.C. Circuit criticized police for entrapping queer men, holding that the testimony of plainclothes decoys was insufficient to justify lewd solicitation arrests unless it was corroborated and requiring courts to give special deference to defendants'

332. *Id*. at 829.

333. 405 U.S. 156 (1972).

334. *Id.* at 162–63, 169–70; JACKSONVILLE, FLA., CODE § 26–27 (1965).

335. *Id.* at 158–59.

336. 349 F.2d 182 (D.C. Cir. 1965).

337. *Id*. at 185. After the Commission tried to explain the grounds for firing Scott, the D.C. Circuit still found it unlawful because the Commission had neglected to give Scott any notice. Scott v. Macy, 402 F.2d 644, 647 (D.C. Cir. 1968).

338. 234 P.2d 969 (Cal. 1951).

339. 222 P.2d 678, 682 (Cal. Dist. Ct. App. 1950) (describing the facts of the case).

340. *Stoumen*, 234 P.2d at 971. Similar decisions in New Jersey and New York in 1967 helped rein in some of the states' worst anti-vice excesses. One Eleven Wines & Liquors v. Div. of Alcoholic Beverage Control, 235 A.2d 12, 19 (N.J. 1967); Becker v. N.Y. State Liquor Auth., 234 N.E.2d 443, 444 (N.Y. 1967).

341. 194 F.2d 150 (D.C. Cir. 1952).

character witnesses.[342] Not all courts stepped in to protect queer victims of overzealous vice squads, and some of the vice squads' worst anti-queer campaigns were the result of a push-and-pull between police and the courts.[343] But at least the courts were a site of contestation about the limits of policing, due process, and equal dignity for queer people.[344]

However, constitutional protections against discrimination, invasions of privacy, insufficient evidence, and police misconduct are not available to queer people disproportionately silenced by sexual content moderation's assemblage of power. Queer users have tried and failed to hold YouTube liable for its disproportionate censoring of queer content.[345] The broad immunity in section 230 of the Communications Decency Act gives platforms wide latitude to experiment with content moderation without liability for the harm they cause in their quest for advertising dollars and "family-friendly" spaces.[346] The state action doctrine bars application of constitutional limits on private corporate behavior.[347] There are no intermediaries with aligned interests that can sue platforms, like bar owners suing the state in the anti-vice era.[348] Platforms have leveraged the First Amendment and its "marketplace of ideas" metaphor to "stave off protective regulation and deflect accountability" for the harm they cause.[349] Progressive prosecutors have no counterpart in today's content moderation; takedown decisions are made by machines and nonexpert humans. And although queer advocacy groups can work with social media companies like they worked with cities and legislatures in the anti-vice era by publishing reports, sitting on advisory trust and safety councils, and putting pressure on platforms to change their sexual content moderation practices, the exogenous mandates only law can provide are

342. *Id.* at 153–56.

343. LVOVSKY, *supra* note 9, at 136–41.

344. *Id.* at 6–7, 134–36, 197–98.

345. *E.g.*, Order Granting Motion to Dismiss, Divino Grp. LLC v. Google LLC, No. 19-cv-04749-VKD (N.D. Cal. Jan. 6, 2021).

346. *E.g.*, Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 588 (S.D.N.Y. 2018).

347. *See* Stephen Gardbaum, *The "Horizontal Effect" of Constitutional Rights*, 102 MICH. L. REV. 387, 388 (2003); Erwin Chemerinsky, *Rethinking State Action*, 80 NW. U. L. REV. 503, 507–08 (1985); *see also* The Civil Rights Cases, 109 U.S. 3, 11 (1883) ("It is [s]tate action of a particular character that is prohibited. Individual invasion of individual rights is not the subject-matter of the [Fourteenth] [A]mendment.").

348. *Stoumen*, *One Eleven Wines*, and *Becker* were all cases brought by bar owners against state regulators who wanted to shut down their establishments. Stoumen v. Reilly, 222 P.2d 678, 680 (Cal. Ct. App. 1950); One Eleven Wines & Liquors, Inc. v. Div. of Alcoholic Beverage Control, 235 A.2d 12, 14–15 (N.J. 1967); Becker v. N.Y. State Liquor Auth., 234 N.E.2d 443 (N.Y. 1967).

349. COHEN, *supra* note 15, at 76, 89–90.

inaccessible.[350] Individuals can and have resisted. But small protests—replacing banned language with emojis, removing tags, and signaling queer content with code to avoid algorithmic curation[351]—are insufficient; platforms have even more power than the state to discriminate against queer, nonnormative expression.

The result is the reification of social media as a straight space. Because of the ways contexts, discourses, rules, technologies, and people interact to enforce sexual norms—interactions explained by the analogy to anti-vice policing of queer sexual behavior—one of social media's distinctive features is the "overwhelming presence" of heterosexual people and the disproportionate "absence" of queer people.[352] Queer people start out with a "deficit of credibility" because their identities, behaviors, and cultural expressions are routinely placed on the wrong side of society's sexual hierarchy.[353] They are allowed in when their expression follows the norms of heterosexual expression.

Queer people are not alone on the outside looking in. As Anderson explains, those inhabiting "[B]lack space" are not only those who identify as Black or African American.[354] Queer space includes all those with nonnormative sexual identities and expressions. And one way these marginalized groups could be welcomed into a traditionally straight space is by using law to upend the systems that created today's sexual content moderation assemblage. The next Part explores the capacity of law to break down the barriers around straight spaces online.

IV. CREATING QUEER INCLUSIVE SEXUAL CONTENT MODERATION

Given the risks to queer expression and the law's complicity in amplifying and entrenching platform power, the legal and regulatory context in which sexual content moderation occurs seems ripe for reform. Achieving that reform in practice will be difficult. There is no magic

---

350. GLAAD, for example, published its *Social Media Safety Index* in 2021 and found that all major social media platforms are unsafe for queer people. *Social Media Safety Index*, GLAAD 6–7 (May 2021). GLAAD, along with queer and queer-friendly organizations like Black Rainbow and the Cyber Civil Rights Initiative, sit on Twitter's Trust and Safety Council. *Trust and Safety Council*, TWITTER, https://about.twitter.com/en/our-priorities/healthy-conversations/trust-and-safety-council [https://perma.cc/H9DR-2E72].

351. *E.g.*, Roth, *supra* note 88, at 427–28; Ysabel Gerrard, *Beyond the Hashtag: Circumventing Content Moderation on Social Media*, 20 NEW MEDIA & SOC'Y 4492, 4498, 4500–02 (2018) (describing evasion of moderation in the pro-eating disorder community).

352. Anderson, *supra* note 14, at 13.

353. *Id.* at 14.

354. *Id.*

bullet, no single piece of legislation or court decision that will solve the problems highlighted above. Any solution should be part of larger structural changes that include robust privacy laws that regulate the advertising-based business model and tough anti-trust enforcement against companies that are too big to care about the needs of marginalized populations. That said, this Part asks: What principles or actions, if any, does the analogy between anti-vice policing and sexual content moderation suggest can help break down barriers to social media for queer content?

A. *The Insufficiency of Current Proposals*

Queer people sit in a double bind with respect to content moderation.[355] Too much sexual content moderation disproportionately censors nonnormative sexual content. On the other hand, too little moderation subjects queer people (not to mention women, persons of color, the disabled, and religious minorities, among others) to hate, harassment, sexual exploitation, and nonconsensual pornography.[356] Less moderation arguably stems from a misguided or intentional commitment to free speech (for some); more moderation stems from platforms designing systems to achieve their economic and institutional goals while accommodating a diverse, global audience.

This double bind makes it difficult to disrupt the ways sexual content moderation maintains social media platforms as straight spaces. All digital spaces must moderate sexual content; the alternative is lawlessness, predation, and harm.[357] But because queer expression is in part defined by its differences and departures from traditional norms about sex,[358] queer people will always be disproportionately censored when rules about speech are premised on those norms. In other words, classifying queer

---

355. A double bind is a situation in which there are no good options. *See* MARTHA CHAMALLAS, INTRODUCTION TO FEMINIST LEGAL THEORY 9–10 (2003); JUDITH HALBERSTAM, FEMALE MASCULINITY 5–6 (1998) (referring to the double bind women face from the need to conform to masculine values to be successful but retain femininity to be socially accepted).

356. DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 227–36 (2014) (arguing that platforms' power to moderate means they have "substantial freedom to decide whether and when to tackle" harms like cyber harassment, nonconsensual pornography, and exploitation).

357. Tarleton Gillespie, *Platforms Are Not Intermediaries*, 2 GEO. L. TECH. REV. 198, 202 (2018) ("[C]ontent moderation is the central service platforms offer. Anyone can make a website on which any user can post anything he or she pleases without rules or guidelines. Such an anarchical website would, in all likelihood, quickly become a cesspool of hate and porn and subsequently be abandoned."). This implies that several politically motivated proposals to treat social media platforms as common carriers or explicitly prohibit them from moderating political expression are more than just unhelpful to queer people. They are also dangerous. *See* Joseph R. Biden v. Knight First Am. Inst., 593 U.S. __, 141 S. Ct. 1220 (mem.), 1221–27 (2021) (Thomas, J., concurring).

358. Berlant & Warner, *supra* note 21, at 346; *see also* ESKRIDGE, GAYLAW, *supra* note 16, at 257.

content as synonymous with adult content and, thus, not suitable for children, members of certain cultures, or a general audience, puts it at risk of censorship.[359] The problem seems intractable, and the analogy between anti-vice policing and sexual content moderation highlights several reasons why current proposals to regulate content moderation are insufficient.

Drawing a more nuanced line between acceptable and unacceptable sexual content would be the logical next step. But even if it were theoretically possible to design a system that was clear and inclusive, its nuance would get lost in the algorithmic and commercial moderation systems necessitated by the massive scale of social media today or platforms would shift tactics. Courts with institutional skills in slow, considered deliberation have trouble drawing a nuanced line.[360] It is difficult to imagine an engineer coding a complex definition into an algorithm or a low-paid commercial moderator making nuanced, sensitive decisions in mere seconds. And if that line is too difficult to draw, platforms would fall back on simpler strategies that have achieved their institutional and corporate goals.[361] A similar thing happened in the anti-vice context. When some judges tried to stop the vice squads' aggressive tactics, police often ignored them, waited for more pliant and sympathetic judges, or switched gears entirely, arresting queer men for another vice after a court put limits on an old one.[362] Integrating more nuanced rules into sexual content moderation seems like a good option, but one that may not be effective in practice.

This raises the question: If moderating speech requires wading into "hard moral and political fights" at scale,[363] would those fights be less tricky if the scale was not so large? That is, would social media platforms be less likely to remain straight spaces if they were smaller, competing with others for users, dollars, and legitimacy? If so, this is a queer case for

---

359. A similar phenomenon was at play when sex workers were taken off platforms after Congress passed legislation to crack down on sex trafficking. Sex work was improperly classified as synonymous with sex trafficking despite the obvious differences. *See* Allow States to Fight Online Trafficking Act of 2017, 18 U.S.C. §§ 1591, 1595, 2421A and 47 U.S.C. § 230; *see also* Kendra Albert, *Five Reflections in Four Years of FOSTA/SESTA*, 40 CARDOZO ARTS & ENT. L.J. __ (forthcoming 2022); Kendra Albert, Emily Armbruster, Elizabeth Brundige, Elizabeth Denning, Kimberly Kim, Lorelei Lee, Lindsey Ruff, Korica Simon & Yueyu Yang, *FOSTA in Legal Context*, 52 COLUM. HUM. RTS. L. REV. 1084, 1088–89 (2021) (arguing that FOSTA may have been intended to protect a vulnerable minority from sexual exploitation but actually harms the vulnerable because of their sexuality); Crystal A. Jackson & Jenny Heineman, *Repeal FOSTA and Decriminalize Sex Work*, 17 CONTEXTS 74 (2018).

360. *See supra* notes 181–182 and accompanying text.

361. COHEN, *supra* note 15, at 6 (arguing that corporate actors make opportunistic use of law and other tools to achieve profit-making goals).

362. LVOVSKY, *supra* note 9, at 130–41.

363. Douek, *Governing*, *supra* note 6, at 775; GILLESPIE, *supra* note 7, at 116.

robust enforcement of anti-trust law against the largest social media companies. But although there are many reasons to pursue greater competition within the information industry,[364] smaller firms are not necessarily more queer-friendly.

It is true that the largest platforms justify restricting queer content by nodding to their need to accommodate global audiences with different norms and views on nonnormative sexual practices. For instance, Instagram justifies enhanced moderation of website links deemed "sexual," many of which are used by queer people, teachers, physicians, and others engaged in sexual health and education, by arguing that it wants "content to be appropriate for our diverse global community."[365] Facebook designs its moderation system to allow "more than [two] billion people to freely express themselves across countries and cultures."[366] The company "recognize[s] that words mean different things or affect people differently depending on their local community, language, or background."[367] YouTube states explicitly that its rules must "be applied to content from around the world."[368] When Instagram removed—and affirmed the removal of—a picture of a shirtless trans man marching in New York City's 2021 "Dyke March," it stated that the photo was removed for violating the platform's policies on "hate speech or symbols. Our guidelines are based on our global community, and some audiences may be sensitive to different things."[369] Equating trans men with "hate speech or symbols" is offensive, but Instagram rationalizes it by nodding to its global audience, much of which, the platform says, is hostile to homosexuality, let alone willing to accept that gender is a social construct. Therefore, to satisfy their conception of their world-wide audience, platforms have an incentive to race to the strictest standards. Unconstrained by free speech rules that limit state censorship, social media companies with multinational reach may find it easier to restrict

---

364. *See* TIM WU, THE CURSE OF BIGNESS 119–27 (2018); Lina Khan, Note, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710 (2017).

365. Abigail Moss, *'Such a Backwards Step': Instagram Is Now Censoring Sex Education Accounts*, VICE (Jan. 8, 2021, 6:56 AM), https://www.vice.com/en/article/y3g58m/instagram-rules-censoring-sex-educators [https://perma.cc/ZYR5-CYDQ].

366. *Facebook Community Standards*, META, https://www.facebook.com/communitystandards/introduction [https://perma.cc/QKV4-7HM8].

367. Monika Bickert, *Updating the Values that Inform Our Community Standards*, META (Sept. 12, 2019), https://about.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/ [https://perma.cc/JEJ3-6UCU].

368. *Rules and Policies*, *Developing Community Guidelines*, YOUTUBE, https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#developing-policies [https://perma.cc/8K9P-MNDL].

369. Renaldi Photos, @renaldiphotos, INSTAGRAM (July 1, 2021), https://www.instagram.com/p/CQynMmIBCk9/?utm_medium=share_sheet (last visited Oct. 13, 2022).

globally the sexual content that is not allowed in the most conservative countries.[370] Using one strict rule everywhere is easier than changing course for different cultures.[371]

Scholars have presumed that greater competition would affect content moderation rules. Writing before the advent of social media, David Post argued that platforms' financial incentives would result in a "market for rules" in which users would be able to seek out platforms with "rule-sets" they prefer.[372] With some amendments, Kate Klonick called Post's view a "useful heuristic to understand the history of online content moderation," adding that the "primary reason" platforms moderate is their "sense of the bottom-line."[373] If a platform takes "down too much content," it loses "not only the opportunity for interaction, but also the potential trust of users."[374] If it takes down too little, it threatens to create an environment that tolerates hate and harassment. At that point, users' expectations would be violated, and they would walk.[375]

Therefore, the conventional wisdom suggests that pursuing robust antitrust enforcement against the information industry's monopolists can only help improve content moderation. This Article has exposed the weakness of that argument. Competition is exogenous to the assemblage of forces at work inside sexual content moderation. Even if platforms no longer had to cater to a global audience with widely divergent views on sex, there remain widely divergent views on sex within individual countries, states, cities, and even households. And smaller platforms' need for advertising dollars would still put sexual content moderation on a path toward disproportionate censorship of queer content. Granted, queer people could flock to a new social media platform specifically designed to be queer-friendly, just like some conservatives started using Parler.[376] But that capacity exists now, and Parler is not immune from moderation

370. Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1039–40 (2018).

371. *E.g.*, Danielle Keats Citron, *The Privacy Policymaking of State Attorneys General*, 92 NOTRE DAME L. REV. 747, 762 (2016) (referring to the relative ease for compliance with a common denominator rule).

372. David G. Post, *Anarchy, State, and the Internet: An Essay on Law-Making in Cyberspace*, 1995 J. ONLINE L., art. 3, ¶ 42.

373. Klonick, *supra* note 7, at 1629, 1627.

374. *Id.* at 1627.

375. *Id.* at 1627–28.

376. Jessica Schulberg, *On Parler, the Right-Wing Social Media Site, Free Speech Isn't Free*, HUFFPOST (June 26, 2020, 7:55 PM), https://www.huffpost.com/entry/parler-free-speech-alternative-twitter-user-agreement_n_5ef660fdc5b6acab28419a5d [https://perma.cc/K579-LERY].

controversy.[377] Plus, forcing queer content into an ancillary corner of the social media ecosystem still denies queer people access to spaces where everyone else is welcome. It leaves in place structural systems of oppression while putting the onus on discriminated minorities themselves to find other options.[378]

B. Lessons from Anti-Vice Policing

If queer expression is to be included in proposals for reform, scholars need to radically rethink their approach to content moderation. Perhaps some of the legal doctrines and cultural shifts that helped rein in the morality police may also help push back on the effects of sexual content moderation's assemblage of power. This Part demonstrates that the anti-vice analogy offers law some options—particularly requiring evidence and robust, meaningful transparency—if policymakers want to guide content moderation in a fairer direction. Let's walk before we run.

Far from monolithic, the law was a site of contestation about the legitimacy of police anti-vice campaigns and the rights of queer people, in general. As Lvovsky argues, some of those debates stemmed from trial courts' "creative intervention" into anti-vice policing "in the face of harsh criminal laws, punitive policies by police departments, and unforgiving legal doctrines imposed by the higher courts."[379] Often, that meant summarily dismissing disorderly conduct, solicitation, assault, and lewdness charges.[380] Elsewhere, courts imposed due process requirements on police departments, rejecting arrests based on uncorroborated testimony from undercover cops or requiring district attorneys, liquor regulators, and police to justify their campaigns with evidence.[381] Indeed, many of the canonical cases of the era turned on the lack of evidence. For instance, in *Scott v. Macy*, a court stopped the Civil Service Commission from firing a queer man because the Commission had not explained how his disorderly conduct arrests made him incapable of doing his job.[382] And

377. Rachel Lerman, *The Conservative Alternative to Twitter Wants to Be a Place for Free Speech for All. It Turns Out, Rules Still Apply*, WASH. POST (July 15, 2020, 10:48 AM), https://www.washingtonpost.com/technology/2020/07/15/parler-conservative-twitter-alternative/ [https://perma.cc/EP5W-ZPBE].

378. *See* Heart of Atlanta Motel v. United States, 379 U.S. 241, 279–284 (1964) (Douglas, J., concurring); *see also* Robert v. U.S. Jaycees, 468 U.S. 609, 625 (1984) (citing *Heart of Atlanta Motel* as a vindication against "the deprivation of personal dignity that surely accompanies denials of equal access to public establishments").

379. LVOVSKY, *supra* note 9, at 7–8.

380. *Id.* at 99, 123, 130, 134, 171.

381. *E.g.*, Kelly v. United States, 194 F.2d 150, 151–56 (D.C. Cir. 1952).

382. 349 F.2d 182, 185 (D.C. Cir. 1965).

Stoumen v. Reilly required police to provide evidence of disorderly conduct before shutting down a bar.[383] To be sure, some of these cases had perverse effects: Police shifted their strategies rather than pulling back on their arrests and, in some cases, subjected their queer arrestees to brutal beatings because they presumed the courts would let them free.[384] But evidentiary requirements placed some limits on police conduct. Therefore, the lesson from the anti-vice analogy is to introduce contestation, evidence, and radical transparency into sexual content moderation.

1. Evidence, not Symbols

Instead of committing to these basic principles of due process, platforms offer false and cynical performances. Some platforms have built endogenous structures to ostensibly provide some measure of contestation and accountability over content moderation. They have hired chief ethics officers to amplify equitable and socially just approaches to AI, including algorithmic moderation.[385] Facebook, YouTube, and other platforms now release regular transparency reports that summarize content moderation actions, takedowns, and restrictions.[386] And Facebook's Oversight Board is an internal council of paid experts that has the power to review some content moderation decisions.[387]

Some scholars want to continue along this path.[388] Along these lines,

---

383. 234 P.2d 969, 971 (Cal. 1951).

384. LVOVSKY, *supra* note 9, at 138–39.

385. *Rise of the Chief Ethics Officer*, FORBES (Mar. 27, 2019), https://www.forbes.com/sites/insights-intelai/2019/03/27/rise-of-the-chief-ethics-officer/?sh=44319f845aba [https://perma.cc/932A-BT3M].

386. *E.g.*, *Transparency, Rules Enforcement*, TWITTER (Jan. 11, 2021), https://transparency.twitter.com/en/reports/rules-enforcement.html#2020-jan-jun [https://perma.cc/8RC4-269D]; *Google Transparency Report*, *YouTube Community Guidelines Enforcement*, GOOGLE (Sept. 28, 2022) https://transparencyreport.google.com/youtube-policy/removals?hl=en [https://perma.cc/2G2X-9PVH?type=image].

387. Klonick, *Oversight*, *supra* note 71, at 2457–74.

388. Douek, *Systems*, *supra* note 13 (manuscript at 62–78). This path is best described as "new" or "collaborative governance," which is an approach to regulation that relies on a partnership between public authorities and private actors to achieve regulatory goals. Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CALIF. L. REV. 1529, 1559 (2019). At its best, new governance is "a highly tailored, site-calibrated regulatory system that aims to pull inputs from, obtain buy-in from, and affect the internal institutional structures and decision-making heuristics of the private sector" while maintaining popular legitimacy and achieving better social welfare outcomes. *Id.* at 1560. For a more comprehensive definition of collaborative governance, see Jody Freeman, *Collaborative Governance in the Administrative State*, 45 UCLA L. REV. 1, 21–33 (1997); Orly Lobel, *New Governance as Regulatory Governance*, *in* THE OXFORD HANDBOOK OF GOVERNANCE 65–67 (David Levi-Faur ed. 2012); Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 371–76 (2004).

Evelyn Douek has proposed ex ante restructuring "content moderation bureaucracies" to mitigate biases.[389] Professor Douek suggests that companies should hire "rule-enforcement personnel" and erect internal separations between those responsible for moderation and those responsible for growth and engagement.[390] They argue for significant transparency so individuals know the nature and extent of the role of outside decision makers in content moderation decisions.[391] And they call for annual compliance reports, auditing, and internal review mechanisms.[392] In other words, internal organizational structures and procedures in the new governance model are supposed to hold platforms accountable for the content moderations they make.[393]

These internal structures are marketed as introducing new voices into the content moderation machinery. But they are insufficient. As Lauren Edelman argued in the Title VII context and I have argued in the privacy context, these types of programs can often become "merely symbolic": They offer the veneer or trappings of legal-adjacent process, but without the substance of real accountability.[394] They do not have counterparts in the anti-vice era; a symbolic structure of accountability would be New York's State Liquor Authority (SLA) setting up a review board chaired by one of its own agents or an internal police review board controlled by the very police officers harassing and arresting queer people for walking down the street. But even if those staffing these positions earnestly cared about accountability, the system is stacked against them. Ethics officers may have little power, with no access to leadership, and serve to deflect attention from corporate actions—like Google's firing of AI ethicist Timnit Gebru—that suggest corporate disinterest in ethical AI.[395] Transparency reports provide some information about the "black box" of content moderation, but they do not explain how or why decisions are made.[396] And despite considerable media and scholarly attention, Facebook's Oversight Board has so little power that it is hard to imagine

389. Douek, *Systems*, *supra* note 13 (manuscript at 61).

390. *Id.* (manuscript at 62).

391. *Id.* (manuscript at 67).

392. *Id.* (manuscript at 69–78).

393. *Id.* (manuscript at 6–7).

394. LAUREN EDELMAN, WORKING LAW 112–13 (2016); Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 778 (2020).

395. Cade Metz & Daisuke Wakabayashi, *Google Researcher Says She Was Fired over Paper Highlighting Bias in A.I.*, N.Y. TIMES (Dec. 3, 2020), https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html [https://perma.cc/8S8Y-CGJW].

396. Douek, *Governing*, *supra* note 6, at 804–08.

it having any real impact on queer expression on the ground.[397] These and other structures on the "periphery of the regulatory state" also sit within corporate organizational hierarchies designed to achieve corporate goals, not limit them.[398] Therefore, it is hard to imagine them suddenly functioning at their best in the content moderation space.

That said, even assuming internal compliance mechanisms do bring some measure of accountability to the administration of content moderation, it is likely that queer expression will be left out. New compliance requirements leave content moderation's inherent heteronormativity untouched. The underlying business model that subjects queer content to the supposedly "common sense" understandings and preferences of advertisers and global markets remains active and, as is likely, further legitimized now that it is framed within mechanisms that have the trappings of law. As a result, they are, at best, performances and, at worst, smoke screens.

Instead of relying on the internal structures and procedures of companies that have shown little to no interest in protecting queer expression, social media websites should be required to provide specific, detailed, case-by-case evidence and explanations for actions they take against content. Individuals should have robust opportunities for contestation, providing evidence and context of their own. I am unmoved by corporate talking points and apologists who claim that content moderation is too complex and too vast to provide the kind of information necessary to demonstrate fairness. Social media companies should not be absolved of basic legal responsibility simply because they are too large. That, as they say, is their problem.

2. *Public Governance, not False Transparency*

Like anti-vice enforcement, sexual content moderation happens without evidence. The process is a black box; users are told their content was removed for violating rules against "sexual activity," but they are not told how, why, or what to fix.[399] Nor are platforms transparent about the interests their moderators are balancing when making decisions.[400] They justify their opacity by arguing that more details about moderation

---

397. Siva Vaidhyanathan, *Facebook Is Pretending It Cares How Its Platform Affects the World*, GUARDIAN (May 6, 2021, 6:24 AM), https://www.theguardian.com/commentisfree/2021/may/06/facebook-donald-trump-ruling-oversight-board, [https://perma.cc/NMQ7-RCL7].

398. COHEN, *supra* note 15, at 192–93.

399. Roth, *supra* note 88, at 415, 418, 423.

400. Douek, *Governing*, *supra* note 6, at 803–04.

practices would allow opportunistic gaming.[401] But, as Sarah Roberts explains, their very reluctance to share information about how sexual content moderation works amplifies platform power to moderate content however they want.[402] Therefore, just as some courts required police to provide legitimate evidence for their arrests, a modest intervention would require platforms to provide more detailed explanations to users when moderating content, allow users to cure, and subject platforms to some measure of legitimate due process.[403]

The anti-vice context also suggests that transparency can do more work to support queer expression on social media. Researchers have always found it difficult to parse arrest records for data about the rates of arrests of queer people versus non-queer people during the anti-vice era.[404] Today, where ethnographic fieldwork suggests that transgender women are overpoliced and harassed on the streets for merely "walking while trans," that problem is even more acute. As Lenore Carpenter and Barrett Marshall note, there is no "police-generated data that could be used to prove more convincingly that transgender women are being subjected to police profiling" because police do not record it, instead relying on forms limited to a gender binary.[405] Carpenter and Marshall argue that sensitively gathering that data could be a first step toward reform.[406]

Similarly, much of the evidence of disproportionate silencing of queer content online is anecdotal, ethnographic, or, in a few circumstances, based on researchers' attempts to recreate the algorithmic moderation environment.[407] Though important and weighty in their own right, those data sources are often viewed with skepticism by platforms, policymakers, courts, and scholars. But quantitative, broad-based, statistical analyses of content moderation's raw data is difficult for one simple reason: platforms hoard that data.[408] They ban scraping in their terms of service and by design.[409] When researchers try gathering data on

---

401. Gerrard, *supra* note 351, at 4493.

402. Roberts, *Detritus*, *supra* note 7 (discussing the logic of opacity and the goals of opaque content moderation).

403. *See* TYLER, *supra* note 254, at 3–10 (arguing that due process enhances police legitimacy).

404. CHAUNCEY, *supra* note 97; Eskridge, *Challenging*, *supra* note 94.

405. Carpenter & Marshall, *supra* note 166, at 6.

406. *Id.* at 22.

407. See studies and evidence described in *supra* section II.B.3.

408. Marco Bastos & Shawn Walker, *Facebook's Data Lockdown Is a Disaster for Academic Researchers*, CONVERSATION (Apr. 11, 2018, 6:10 AM), https://theconversation.com/facebooks-data-lockdown-is-a-disaster-for-academic-researchers-94533 [https://perma.cc/3WC7-N3H2].

409. COHEN, *supra* note 15, at 235.

their own, platforms sue to stop them.[410] They permit research by some researchers and not others on their terms.[411] As Amy Kapczynski has noted, the combination of contractual and legal limits on researcher access to data allows platforms to "forbid users from undertaking research that might disclose aspects of their platform's functioning" and hold them accountable for abuses of power.[412]

Therefore, as in the "walking while trans" context, where advocates have long called for data gathering on queer populations to prevent systemic erasure, another proposal for legal intervention would require platforms to make content moderation data available to researchers. Scholars could then interrogate, analyze, and study under circumstances that both protect platforms' trade secrets and permit independent interrogation. That data can include what is taken down or restricted and why, which posts are flagged and how they are reviewed, and which content is algorithmically limited. If platforms are indeed doing their best to balance competing interests, such transparency could also provide a legitimacy dividend. Notably, here I agree with Douek and other scholars who deftly call for robust transparency and legal protection for transparency.[413]

With this data, independent researchers—not people paid by platforms themselves—can play the role that some judges played in the anti-vice context and introduce contestation into the content moderation process. They can assess the veracity of platforms' commitments to fair moderation, hold platforms to their promises, and catalyze change by publicizing their research. There is a model for this. Julia Angwin and other investigative journalists have changed the discourse around algorithmic bias and published research proving that platforms have failed to keep their commitments.[414] Researchers like Joy Buolamwini, Safiya

410. *E.g.*, Jeff Horwitz, *Facebook Seeks Shutdown of NYU Research Project into Political Ad Targeting*, WALL ST. J. (Oct. 23, 2020, 8:59 PM), https://www.wsj.com/Articles/facebook-seeks-shutdown-of-nyu-research-project-into-political-ad-targeting-11603488533 (last visited Nov. 18, 2022).

411. NATHANIEL PERSILY & JOSHUA A. TUCKER, *The Challenges and Opportunities for Social Media Research*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM 324 (Nathaniel Persily & Joshua A. Tucker eds., 2020).

412. Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1460, 1502–03 (2020). *See also* COHEN, *supra* note 15, at 116.

413. Douek, *Systems*, *supra* note 13 (manuscript at 68–69).

414. *E.g.*, Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/Article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/VF2P-8JGF] (identifying racial bias in recidivism algorithms); Corin Faife & Alfred Ng, *After Repeatedly Promising Not To, Facebook Keeps Recommending Political Groups to Its Users*, MARKUP (June 24, 2021, 8:00 AM), https://themarkup.org/citizen-browser/2021/06/24/after-

Noble, and Latanya Sweeney have used technical tools to highlight algorithmic biases.[415] If they had the tools, these and other sociotechnical researchers can act as a counterweight to platform power, prying open content moderation's "black box" and holding social media accountable.

3. *Anti-Subordination, not Blanket Immunity*

This kind of transparency and research also means that more significant legal reform is possible. Danielle Citron, Mary Anne Franks, and other scholars have argued for reforming the CDA section 230 immunity to better incentivize platforms to restrict hateful, harassing, and illegal content.[416] Their proposals would amend section 230 to condition immunity on platforms "engag[ing] in good faith efforts to restrict illegal activity."[417] Franks would limit immunity only to "speech wholly provided" by a user unless the platform "intentionally encourages, solicits, or generates revenue" from the offending speech.[418] If researchers had content moderation data available to them that could demonstrate or rule out disproportionate censorship of queer content, section 230 could also be amended to condition immunity on platforms engaging in good faith efforts to ensure that their policies and practices do not result in disparate impact—namely, disproportionate silencing or throttling—on those voices marginalized under traditional norms. Those efforts would be documented by platforms and subject to public governance oversight, not internal compliance mechanisms.

I would also go further and propose an anti-subordination element to legal reform. Anti-subordination refers to the idea that equality doctrine should not simply prohibit classifications on the basis of demographic criteria, but that "lived equality—that is, substantive, material, day-to-day equality as opposed to formal, 'on-the-books' equality—necessitates

---

repeatedly-promising-not-to-facebook-keeps-recommending-political-groups-to-its-users [https://perma.cc/8TWD-9JRS].

415. *E.g.*, SAFIYA NOBLE, ALGORITHMS OF OPPRESSION (2018); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. OF MACHINE LEARNING RES. 1 (2015); Latanya Sweeney, Discrimination in Online Ad Delivery (Jan. 28, 2013) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2208240 [https://perma.cc/9P83-CPVC].

416. *E.g.*, Citron & Wittes, *supra* note 37, at 412–16; Mary Anne Franks, *Sexual Harassment 2.0*, 71 MD. L. REV. 655 (2012); Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. L. F. 45 (2020); MARY ANNE FRANKS, STAN. CYBER POL'Y CTR., REFORMING SECTION 230 AND PLATFORM LIABILITY (2021) [hereinafter FRANKS, REFORMING SECTION 230].

417. Citron & Wittes, *supra* note 37, at 416; *see also* Danielle Keats Citron, *How to Fix Section 230*, 103 B.U. L. REV. __ (forthcoming 2023).

418. FRANKS, REFORMING SECTION 230, *supra* note 416, at 11.

dismantling facially neutral . . . [systems] that nevertheless oppress particular groups."[419] Anti-subordination sometimes requires rule makers to be conscious of certain classifications and affirmatively "level up those that are being subordinated."[420] In other words, unlike today's anti-political moderation proposals from the political right,[421] an anti-subordination agenda to reform section 230 would require platforms to make a concerted effort to understand the disproportionate effects of its moderation systems on content from queer and other oppressed minorities and make design and policy changes to ameliorate them systemically. That is, unlike managerialized compliance mechanisms, which tend to serve corporate goals like efficiency and profit while reifying existing business models, exogenous legal requirements that level up marginalized populations could have the salutary effect of changing the very nature of digital social spaces. And that structural change is necessary if queer expression ever hopes to be a robust part of digital life.

Ultimately, though, the anti-vice analogy teaches us that no legal intervention can replace the community activism and cultural shifts that made some lawyers and judges change their minds about morality policing. As Lvosvky notes, where "social conservativism and conventional ideals of masculinity" among police fomented a natural antipathy toward queer people, more judges in the 1960s were influenced by changing social mores around sex, gender, and homosexuality.[422] Civil rights and sexual liberation movements, together with broader progressive political and medical movements, started a slow shift in discourse about the dignity of queer men.[423] Homophile organizations like the Mattachine Society and One also aggressively lobbied legal advocacy groups like the ACLU to encourage the group's attorneys to take up the cause of the mostly queer men being victimized by morality policing.[424] Simply put, times changed and the assemblage of forces that turned anti-vice policing into a weapon against queer people shifted, evolved, and weakened. A similar shift in culture about sex, gender, and sexuality may be queer content's best hope in the end.

---

419. SCOTT SKINNER-THOMPSON, PRIVACY AT THE MARGINS 6 (2020).

420. *Id.*

421. *See* NetChoice v. Paxton, No. 21-51178, 2022 WL 4285917 (5th Cir. Sept. 16, 2022) (upholding Texas law that limited content moderation of political viewpoints); Ending Support for Internet Censorship Act, S. 1914, 116th Cong. (2019); Russell Brandom, *Senate Republicans Want to Make It Easier to Sue Tech Companies for Bias*, VERGE (June 17, 2020, 6:46 AM), https://www.theverge.com/2020/6/17/21294032/section-230-hawley-rubio-conservative-bias-lawsuit-good-faith [https://perma.cc/EM9V-8M4N].

422. LVOVSKY, *supra* note 9, at 118–19, 139.

423. *Id.* at 118.

424. *Id.* at 174–75.

CONCLUSION

Through a historical analogy, this Article has shown that sexual content moderation systemically encodes queerness as illicit in ways straightness is not. Sexual content moderation operates like and has similar effects on public expressions of queer sexuality as anti-vice policing from the 1930s to the late 1960s. Both are assemblages of social forces comprising law, ideas, rules, technologies, expertise, and people. And both have the effect of creating and maintaining straight spaces. Although few historical analogies are perfect, and it is important not to equate policing and subjugation by the state with policing sexuality by a private company, the parallels highlight several critical lessons for content moderation, technology, and law. The analogy also offers new perspective on current policy debates about platform regulation.

Queer people are certainly not the only ones who receive disparate treatment by social media. But the values, machinery, and anti-queer effects of sexual content moderation are one example in a larger narrative about technologies' role in queer oppression and the reification of social media as a straight space. Platforms that play critical roles in the production and dissemination of knowledge also control spaces characterized by harassment and misinformation, both of which endanger queer lives. Together, anti-queer censorship, attacks on queer identity, and the dissemination of lies and stereotypes about queer people deny queer access to social technologies and ultimately amplify discourses of marginalization and discrimination.

This Article's analogy to anti-vice policing has shown just how much the deck is stacked against queer expression in today's social media landscape even when those platforms are not designed with express and malicious anti-queer intentions. Therefore, platforms have a choice: They can maintain a discriminatory status quo *or* they can commit to anti-subordination. Law has a choice, too: To focus on procedural guardrails is to endorse the disproportionate silencing of marginalized voices; to force platforms down to size, to commit them to protecting the marginalized, to condition immunity on anti-subordination is to fulfill the state's commitment to its most vulnerable. To step in may be fraught, but it is the only path forward.