2020

Bad Actors: Authenticity, Inauthenticity, Speech, and Capitalism

Sarah C. Haan
Washington and Lee University School of Law, haans@wlu.edu

ARTICLES

BAD ACTORS: AUTHENTICITY, INAUTHENTICITY, SPEECH, AND CAPITALISM

*Sarah C. Haan**

ABSTRACT

"Authenticity" has evolved into an important value that guides social media companies' regulation of online speech. It is enforced through rules and practices that include real-name policies, Terms of Service requiring users to present only accurate information about themselves, community guidelines that prohibit "coordinated inauthentic behavior," verification practices, product features, and more.

This Article critically examines authenticity regulation by the social media industry, including companies' claims that authenticity is a moral virtue, an expressive value, and a pragmatic necessity for online communication. It explains how authenticity regulation provides economic value to companies engaged in "information capitalism," "data capitalism," and "surveillance capitalism." It also explores how companies' self-regulatory focus on authenticity shapes users' views about objectionable speech, upends traditional commitments to pseudonymous political expression, and encourages collaboration between the State and private companies. The Article concludes that "authenticity," as conceptualized by the industry, is not an important value for users on par with privacy or dignity, but that it offers business value to companies. Authenticity regulation also provides many of the same opportunities for viewpoint discrimination as does garden-variety content moderation.

* Associate Professor of Law, Washington and Lee University School of Law. The Author thanks Carliss Chatman, Rebecca Green, Margaret Hu, Lyman Johnson, Thomas Kadri, James Nelson, Elizabeth Pollman, Carla L. Reyes, Christopher B. Seaman, Micah Schwartzman, Morgan Weiland, participants in the W&L Law 2019 Big Data Research Colloquium, and participants in the 2019 Yale Freedom of Expression Scholars Conference ("FESC VII"), for their insightful comments on early drafts of this Article. She also gratefully acknowledges that funding from the Frances Lewis Law Center supported this work. In addition, the Author wishes to disclose that she owns a small amount of stock in Facebook, Inc., and operates a Twitter account at @shaan_haan.

TABLE OF CONTENTS

INTRODUCTION

In 2015 and 2016, Russian-linked groups ran paid content on Facebook in an effort to influence the U.S. election.[1] When Facebook publicly acknowledged this in September 2017, the company was careful in its framing. The *content* of the offending advertisements was not a problem, Facebook's executives explained.[2] Rather, the problem was the "inauthenticity" of their sources; the Russians were *bad actors* because they had pretended to be someone they were not.[3] Sheryl Sandberg, Facebook's Chief Operating Officer, told an interviewer that most of the Russian-linked advertisements would have been permitted on Facebook "if they were run by legitimate people," meaning people who presented their true identities.[4]

The company's choice of framing was significant. When Congress enacted laws criminalizing foreign election interference, speaker deception was not the problem it sought to address. Federal law prohibits foreign interference regardless of whether the speaker presents a true or false identity, on the view that foreign influence distorts the political process even when it is undisguised.[5] The social media industry, on the other hand, has

---

1 *See* Alex Stamos, *An Update on Information Operations on Facebook*, FACEBOOK NEWSROOM (Sept. 6, 2017), https://newsroom.fb.com/news/2017/09/information-operations-update/ (explaining how Facebook identified thousands of dollars in advertisements purchased by inauthentic Russian-linked pages).

2 *See Hearing Before the Subcomm. on Crime & Terrorism of the S. Comm. on the Judiciary*, 115th Cong. 5 (2017) (statement of Colin Stretch, General Counsel, Facebook), *available at* https://www.judiciary.senate.gov/imo/media/doc/10-31-17%20Stretch%20Testimony.pdf ("The Facebook accounts that appeared tied to the IRA violated our policies because they came from a set of coordinated, inauthentic accounts.") [hereinafter Testimony of Colin Stretch]; Elliot Schrage, *Hard Questions: Russian Ads Delivered to Congress*, FACEBOOK NEWSROOM (Oct. 2, 2017), https://newsroom.fb.com/news/2017/10/hard-questions-russian-ads-delivered-to-congress/ ("We require authenticity regardless of location. If Americans conducted a coordinated, inauthentic operation—as the Russian organization did in this case—we would take their ads down too. However, many of these ads did not violate our content policies. That means that for most of them, if they had been run by authentic individuals, anywhere, they could have remained on the platform.").

3 Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/ (describing the Internet Research Agency as a "bad actor[]").

4 *Exclusive Interview with Facebook's Sheryl Sandberg*, AXIOS (Oct. 17, 2017), https://www.axios.com/exclusive-interview-with-facebooks-sheryl-sandberg-1513306121-64e900b7-55da-4087-afee-92713cbbfa81.html (reiterating points from Elliot Schrage's earlier blog post); *see also* Schrage, *supra* note 2 (noting Facebook's talking points in response to the Russia election interference inquiry).

5 52 U.S.C. § 30121 (2018); *See* Bluman v. FEC, 800 F. Supp. 2d 281, 288 (D.D.C. 2011) (exclusion of foreign citizens from activities of democratic self-government is necessary to preserve "our national political community"), *aff'd* 565 U.S. 1104 (2012); Zephyr Teachout, *Extraterritorial Electioneering and the Globalization of American Elections*, 27 BERKELEY J. INT'L LAW 162, 183–187 (2009)

long enforced private authenticity rules. Companies require users to present only their "true" selves on social media, and censor "inauthentic" speech.[6] As this Article documents, authenticity enforcement is expanding, with few critics. Over roughly a decade, authenticity has evolved into an important "value" used to shape online speech.[7] What accounts for the rise of authenticity? The companies argue that authenticity is a quality of personal integrity, a cudgel to reduce abusive behavior and crime, and an essential component of free expression. They also suggest that authentic speakers produce *authentic content*. This Article, which critically examines authenticity regulation by private companies, explores additional reasons that are not commonly discussed, including that companies engaged in "information capitalism,"[8] "data capitalism,"[9] and "surveillance capitalism"[10] derive economic value from authenticity regulation.[11]

Free-speech jurisprudence recognizes that the State can burden speech through many regulatory methods. One method is content-based regulation, in which the State singles out some speech for special treatment, or outlaws it altogether, based on the substance of what it communicates.[12] Another method is speaker-based regulation, in which the State treats speech differently based upon who is speaking.[13] In either case, a main concern is

---

(discussing the self-government and sovereignty interests that have traditionally justified a prohibition against foreign election interference).

6 *See infra* Part I.B.

7 *See, e.g.*, *Facebook Community Standards*, FACEBOOK, https://www.facebook.com/community standards/ (discussing the five core "values" that shape Facebook's regulation of speech: voice, authenticity, safety, privacy, and dignity) (last visited Jan. 4, 2020).

8 *See generally* Jeremy K. Kessler & David E. Pozen, *The Search for an Egalitarian First Amendment*, 118 COLUM. L. REV. 1953 (2018); Julie Cohen, *The Regulatory State in the Information Age*, 17 THEORETICAL INQUIRIES L. 369 (2016).

9 Jack M. Balkin, *Fixing Social Media's Grand Bargain* 3 (Hoover Institution, Aegis Series Paper No. 1814, 2018), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3266942 (describing "data capitalism" as "the grand bargain of the Second Gilded Age").

10 SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 8–9 (2019).

11 These terms describe an emerging economy of business transactions in which value is extracted from individuals' data through data analytics.

12 *See, e.g.*, Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 640–643 (1994) ("As a general rule, laws that by their terms distinguish favored speech from disfavored speech on the basis of the ideas or views expressed are content based."); Police Dep't of Chi. v. Mosley, 408 U.S. 92, 96 (1972) ("[A]bove all else, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.").

13 Rosenberger v. Rector & Visitors of the Univ. of Va., 515 U.S. 819, 827 (1995) ("The government must abstain from regulating speech when the specific motivating ideology or the opinion or perspective of the speaker is the rationale for the restriction."). A third method is compelled speech, where the State forces a speaker to make a disclosure. Other methods exist as well.

that the State will misuse its power to burden speech with which it disagrees, in order to suppress a particular idea or to manipulate public debate.[14]

Although the First Amendment does not apply to private social media companies, these "New Governors" of speech[15] regulate public discourse online, where they employ the same speech-regulating methods used by state actors, including content-based measures, speaker-based measures, and even mandatory disclosures. Academic study of social media companies' speech regulation has created a rich literature on content moderation, but has been slow to examine other regulatory methods.[16] This Article starts from the proposition that authenticity rules constitute a form of speaker-based speech regulation, because they treat "authentic" speakers differently from "inauthentic" speakers.[17]

The social media network is the first significant speech forum in which a single "curator" controls both the production and receipt of speech by individual participants, determining simultaneously and continuously what each participant is allowed to say and to whom, and what speech each participant is allowed to receive, and how. Even this description fails to capture the full speech-regulating power of the social media user interface, which can be designed to up-rank or down-rank speech relative to other speech, to "push" messages across devices, to repeat messages or deliver them at a particular moment, to make some content more visually engaging than

---

14 *See, e.g.*, Citizens United v. FEC, 558 U.S. 310, 340 (2010) ("Speech restrictions based on the identity of the speaker are all too often simply a means to control content."); *Turner Broad. Sys.*, 512 U.S. at 641 ("Government action that stifles speech on account of its message," i.e., content-based regulation, poses "the inherent risk that the Government seeks . . . to suppress unpopular ideas or information or manipulate the public debate through coercion rather than persuasion.").

15 *See generally* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018) (describing the increasing role and responsibility of private online platforms in free speech and democratic culture).

16 *See generally id.*; Evelyn Aswad, *The Future of Freedom of Expression Online*, 17 DUKE L. & TECH. REV. 26 (2018); Thomas E. Kadri & Kate Klonick, *Facebook v.* Sullivan*: Public Figures and Newsworthiness in Online Speech*, 93 S. CAL. L. REV. 37 (2019); Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353 (2018). Separate literature has looked at problems relating to privacy and discrimination in the use of artificial intelligence, including situations in which artificial intelligence shapes speech and debate on social media. *See, e.g.*, Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. Rev. 54, 59 (2019) ("An algorithm can instantly lead to massive discrimination between groups."); Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMEND. INST. (Apr. 1, 2018), http://knightcolumbia.org/content/discriminatory-designs-user-data.

17 *See Citizens United*, 558 U.S. at 340 (finding speaker-based discrimination occurs when "restrictions distinguish[] among different speakers, allowing speech by some but not others").

others, and to exploit a limitless set of behavioral insights[18] designed to influence the recipient's response to the message. Social media companies like Facebook and Twitter monetize this technological capability by marrying it with a constant inflow of user-specific data that works ceaselessly to identify, distinguish and quantify people for the purpose of determining their speaking and listening prerogatives, and for fixing the fees the companies will charge speakers.

The social media network is not the "marketplace of ideas" imagined by twentieth-century visionaries, in which demand for the best ideas causes them to rise to the top.[19] In the social media exchange, speech is "served" to a group of recipients based upon the amount the speaker is willing (or able) to pay and the recipients' identifying characteristics and behavior. At profit-seeking social media companies, the relationship of these factors is expressed in the form of a proprietary algorithm that has a purpose to maximize payments to a third party: the social media company.

It was a short leap from authenticity enforcement to identity-verification requirements for speakers. After experimenting with verification for public figures for years, in 2017, Facebook rolled out a system of speech licensing for any user who wants to discuss "national issues of public importance" with the use of its paid tools.[20] Introduced as a solution to political advertisement transparency problems, the new rules required speakers to send the company an image of his or her U.S. passport or driver's license so that Facebook could verify the speaker's identity.[21] Verification threatens to exclude certain kinds of people from participation in public discourse online, such as low-income and undocumented individuals. The *Washington Post* found that verification requirements have placed extra burdens on speech that touches on LGBTQ issues, because Facebook has treated all LGBTQ-related content as political speech.[22] Facebook has deployed verification selectively: for a time, under

18 *See, e.g.*, Ricardo Baeza-Yates, *Bias on the Web*, 61 COMMS. ACM 54, 59 (2018) (describing "position bias," in which content that appears in the top left corner of a screen receives a more significant audience response).

19 *See* Annemarie Bridy, *Remediating Social Media: A Layer-Conscious Approach*, 24 B.U. J. SCI. & TECH. L. 193, 217 (2018) ("That process of truth-finding through truth-testing," captured by the marketplace of ideas metaphor, "bears little resemblance to the algorithmic sorting that creates winners and losers in social media's attention sweepstakes.").

20 *See Ads About Social Issues, Elections or Politics*, FACEBOOK AD HELP CTR., https://www.facebook.com/business/help/214754279118974 (last visited Mar. 26, 2020).

21 *See generally id.*

22 *See infra* Part I.C.

its rules, "poverty" was a political issue requiring speakers to verify their identities, but "wealth" was not.[23]

At Twitter, only verified accounts are eligible for exemptions from the company's content rules on "public interest" grounds.[24] And speakers who violate Twitter's content rules have had their identity verification—their blue check mark—rescinded as punishment.[25]

Companies also employ product features that capitalize on users' "true" identities. Facebook has marketed "Town Hall" features that allow elected officials to communicate with tailored audiences comprised only of individuals identified by Facebook as living within the official's area of representation.[26] For example, the feature has allowed elected officials to host virtual town halls on Facebook Live, attended only by verified constituents.[27] These features not only limit who can speak and listen to elected officials, but also which journalists can report on those communications. They purport to define the group of individuals who are *authentic constituents* of an elected official.[28]

Social media companies originally characterized authentic identity as a status, but have re-characterized it over time to include behavior.[29] Both Facebook and Twitter prohibit something they call "coordinated inauthentic behavior."[30] Under this behavioral approach, authentic speakers can violate a company's authenticity rules by behaving in inauthentic ways or by associating with "bad actors." This shift, which has led companies to use

---

23 *See Ads About Social Issues, Elections or Politics*, FACEBOOK AD HELP CTR., https://www.facebook.com/business/help/214754279118974 (last visited Mar. 26, 2020).

24 *See infra* Part I.C.

25 *See infra* Part II.E.

26 Griffin Connolly, *Facebook Features Connect Lawmakers with Constituents*, ROLL CALL (June 8, 2017), https://www.rollcall.com/politics/facebook-features-connect-lawmakers-constituents.

27 *See* Faine Greenwood, *A Civics Lesson for Facebook*, SLATE (Aug. 8, 2017), https://slate.com/technology/2017/08/facebook-now-offers-constituent-services-what-could-go-wrong.html (explaining how Facebook's "Town Hall" project, which included the constituent badge feature, also introduced "district targeting," which allows elected officials to create posts and polls that are visible only to confirmed constituents, and "constituent insights," which provides elected officials with tools to view and comment on news stories that are popular among their constituents).

28 In doing so, they operationalize the company's view on the politically contested issue of who is a "constituent" of an elected official. *See infra* Part I.F. *See generally* Richard Briffault, *Of Constituents and Contributors*, 2015 U. CHI. LEGAL F. 29 (2015) (discussing Chief Justice Roberts's endorsement of "contributor representation" in *McCutcheon v. FEC*, 572 U.S. 185 (2014)).

29 Shoshana Zuboff has argued that the ultimate purpose of surveillance capitalism is to shape or manipulate individuals' behavior. *See generally* ZUBOFF, *supra* note 10; *see also* Balkin, *supra* note 9 (noting how the digital age "exacerbates the twentieth-century problem of manipulation").

30 *See* discussion *infra* Part II.B.4.

machine learning to proactively flag and punish problematic *associations*, potentially burdens individuals' freedom of association, both online and in the real world. In at least one recent case, Facebook identified "coordinated inauthentic behavior" where political liberals, operating under their real names, ran a Page called "Conservative Alabama Politics" that sought to influence conservative voters.[31] The speakers did not misrepresent their identities, but their effort to address a conservative audience was treated as *inauthentic behavior* because they were not genuinely conservative. Twitter has been charged with erroneously sweeping up "real" speakers in purges of networks accused of coordinated inauthentic activity.[32]

Many proposed solutions to social media speech-harms are designed to address concerns about content moderation and privacy, while leaving in place the authenticity rules and practices that give value to data analytics. For example, commentators have called for expansions of the "state actor" doctrine to make it harder for technology companies to engage in content moderation,[33] for the treatment of social media companies as "public utilities,"[34] or for tougher new privacy laws.[35] None of these proposals addresses authenticity regulation, continuing a trend in which commentators and scholars tend not to recognize authenticity policing as a form of speech regulation.

Facebook itself has convened an independent body—a sort of private supreme court—to oversee its content moderation.[36] The Charter for its

---

31 *See infra* notes 123–26 and accompanying text.

32 *See infra* notes 116–18 and accompanying text.

33 *See, e.g.*, Colby M. Everett, *Free Speech on Privately-Owned Fora: A Discussion on Speech Freedoms and Policy for Social Media*, 28 KAN. J.L. & PUB. POL'Y 113, 115 (2018) ("[T]his article argues social media are public fora regulated by quasi-governmental actors seeking to filter certain speech."); Benjamin F. Jackson, *Censorship and Freedom of Expression in the Age of Facebook*, 44 N.M. L. REV. 121, 121–22 (2014) (arguing courts should "deem censorial acts by social network websites to be state action under the public function exception to the state action doctrine"). *But see* Jack M. Balkin, *supra* note 9 (opposing this proposed solution, noting that "social media sites might want to require that end users use their real names or easily identifiable pseudonyms in order to limit trolling and abuse").

34 *See generally* Adam D. Thierer, *The Perils of Classifying Social Media Platforms as Public Utilities*, 21 COMM. L. CONSPECTUS 2 (2013) (discussing the possibility of conferring public utility status on major social media platforms).

35 Jack Balkin has proposed that social media companies and other online service providers be treated as "information fiduciaries" toward customers and other end-users. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1186 (2016).

36 *See* Brent Harris, *Establishing Structure and Governance for an Independent Oversight Board*, FACEBOOK NEWSROOM (Sept. 17, 2019), https://newsroom.fb.com/news/2019/09/oversight-board-structure/ (describing and evaluating the proposal for Facebook's "Supreme Court"); Thomas E. Kadri & Kate Klonick, Facebook v. Sullivan*: Building Constitutional Law for Online Speech*, 93 S. CAL. L. REV. 37, 74–80 (2019).

Oversight Board mentions authenticity once, characterizing it as a potential *limitation* on free expression, rather than as purely beneficial for speech.[37] The company's policy documents present authenticity as one of five values—the others are voice, safety, privacy, and dignity—that the Oversight Board will use "to inform its decisions" on content.[38]

The Article proceeds in two parts. Part I describes how authenticity regulation has evolved in the social media industry, with an emphasis on exploring authenticity as a business value. Relying mainly on the example of Facebook, it reviews the advertisement-based business model, with its reliance on micro-targeting and customization, and shows how authenticity rules underwrite the surveillance-capitalism business model. As it shows, both authenticity policies and identity-verification systems have deepened and expanded in the wake of the Russian-interference scandal, but neither was created in response to that scandal. Rather, these practices continue long-standing strategies that tend to enhance the industry's profit-generating activities.

Companies' authenticity policies make it possible for them to quantify users, to offer paying customers a reliable count of the people who receive their advertisements, and to offer investors a measure of the company's user base, growth, and future cash flows. They also allow companies to ensure the integrity of their user-specific data, which is critical for machine learning. In order for companies' data systems to "learn" patterns of human behavior, they must have accurate data inputs. Thus, a Facebook user who misrepresents his age to Facebook corrupts the company's machine learning, because the system will attribute all of his behaviors to a younger (or older) person and glean false insights about human behavior from that attribution.

Part II assesses authenticity as a core social media value. Companies have variously described authenticity as a moral virtue, a pragmatic necessity, an essential component of "meaningful" speech, and a limit on free expression. They claim that forcing users to present their "true" identities cuts down on harassment and other speech harms. They contend that authenticity-based takedowns are necessary to prevent fraud and other crimes, and to nurture trust in online communications. And they commonly

---

37 *See* FACEBOOK, OVERSIGHT BOARD CHARTER 2 (2019), *available at* https://fbnewsroomus.files.wordpress.com/2019/09/oversight_board_charter.pdf [hereinafter OVERSIGHT BOARD CHARTER]; *see also infra* notes 186–91 and accompanying text.

38 *See generally* MARK ZUCKERBERG, FACEBOOK'S COMMITMENT TO THE OVERSIGHT BOARD (2019), *available at* https://about.fb.com/wp-content/uploads/2019/09/letter-from-mark-zuckerberg-on-oversight-board-charter.pdf (describing the policy and decision behind creating an Oversight Board to support the right to free expression).

elide the difference between authentic *identity* and authentic *content*, suggesting that people who present only true information about themselves produce authentic (i.e., good) speech.

Part II argues that truthful presentation of self *can* have salutary effects on certain kinds of online expression, functioning as a sort of proxy for *truth*. However, it argues that authenticity in the industry sense is *not* a value on par with human rights like privacy and dignity. Part II explores a number of reasons to be concerned about authenticity regulation. By conflating "false identity" with "anonymous identity," it undermines American free speech values that have traditionally protected pseudonymous speakers and anonymous political speech. Authenticity regulation conveys the value judgment that when speech is objectionable, it is because of the identity of the person speaking, and not because of the content of the speech. This value judgment differs from traditional notions of free speech, which acknowledge that some kinds of speech are both objectionable *and* protected from censorship. The evidence of a connection between authenticity and abuse is mixed; some recent research has found that speakers operating under their real names are *more* likely to behave abusively.[39]

When they act as arbiters of authenticity, social media companies enjoy the same power to suppress viewpoints and manipulate debate as they would if they were regulating content. By making "authentic" identity a valuable commodity, companies encourage identity theft, because a stolen identity can be harder to detect as false. In doing so, they encourage an arms-race between technology companies and sophisticated identity thieves, including foreign nation-states. This not only increases the value of identity-verification services, creating profit opportunities for the same technology companies that contributed to the problem, but pushes companies to form reciprocal relationships with law enforcement. On balance, authenticity regulation may make us worse off, providing little "value" for all of our cooperation.

I. THE BUSINESS OF AUTHENTICITY

Facebook's authenticity rules trace their origin to the company's earliest days, in 2004, when a young Mark Zuckerberg conceived its real-name

39 *See infra* Part II.A.3.

policy.[40] David Kirkpatrick, who interviewed Zuckerberg for his 2010 book, wrote that Facebook's real-name policy emerged from Zuckerberg's own "strength of conviction" that the transparency of the Internet required participants to present only one, true self.[41] "Having two identities for yourself," Zuckerberg told Kirkpatrick at the time, "is an example of a lack of integrity."[42] Today, Facebook's rules on authentic identity are found in its Community Standards under the heading "Authenticity and Integrity," underscoring not only the company's presentation of identity as an issue of user morality, but also the continuing influence of Zuckerberg—the company's controlling shareholder, Chief Executive Officer, and board chair—on Facebook's approach to speech regulation.[43]

Facebook sells advertisements, and this requires it to have accurate metrics about who is viewing advertisements on its network. Thus, one purpose of Facebook's authenticity rules is to make the quantification of advertisement recipients easy and accurate: users are forbidden from sharing accounts and human users are carefully distinguished from organizational users through Facebook's profile/Page distinction. In addition, Facebook's advertisement-based business model relies heavily on micro-targeting through data analytics, and this requires the company to maintain a detailed, accurate profile on each user. Facebook's authenticity rules facilitate the

---

40 DAVID KIRKPATRICK, THE FACEBOOK EFFECT 31 (2010); Tom Huddleston, Jr., *Here's How 19-year-old Mark Zuckerberg Described 'The Facebook' in His First TV interview*, CNBC (Apr. 17, 2018), https://www.cnbc.com/2018/04/16/how-mark-zuckerberg-described-the-facebook-in-his-first-tv-interview.html. The orwellian term "authentic identity" did not become part of Facebook's rulebook—its Community Standards—until 2015. *See* Tarleton Gillespie, *Facebook's Improved "Community Standards" Still Can't Resolve the Central Paradox*, SOCIAL MEDIA COLLECTIVE (Mar. 18, 2015), https://socialmediacollective.org/2015/03/18/facebooks-improved-community-standards-still-cant-resolve-the-central-paradox/ (discussing the shift in standards for requiring users to portray themselves accurately).

41 Kirkpatrick described Zuckerberg repeating "You have one identity" three times in a single minute in a 2009 interview and attributed to Zuckerberg both a "moralistic[]" and a "pragmatic" belief that users must present only their true identity on the platform. DAVID KIRKPATRICK, THE FACEBOOK EFFECT 199–200 (2010).

42 This undated quote was attributed to Zuckerberg in David Kirkpatrick's 2010 book, *The Facebook Effect*. *Id.* at 199. See *infra* note 196 and accompanying text for further discussion of why Zuckerberg might have felt moral zeal for identity policing, in light of recent social psychology research.

43 Facebook also sometimes asserts that its real-name policy leads to better user behavior. Recent research has called this conventional wisdom into question. *See infra* Part III.A.3. As danah boyd observed in 2012, "[m]any people claim people are better behaved and more honest when their identifying information is available. While there is no data that convincingly supports or refutes this, it is important to note that both Facebook and face-to-face settings continue to be rife with meanness and cruelty." danah boyd, *The Politics of Real Names*, 55 COMMS. ACM 29, 30 (2012).

company's gathering of accurate, identifying information about each user, which can be matched to user-specific information from other, commercial sources.[44] For example, under the company's Terms of Service, users expressly agree to provide only "accurate" information about themselves to Facebook.[45]

Facebook's authenticity regulation has expanded far beyond the original real-name policy and today involves at least four parts: a set of authentic/inauthentic identity rules and practices, which determine who is allowed to use Facebook's network—i.e., who is able to produce and receive speech; an identity-verification system, which allows or requires certain speakers to verify their identities with the company by submitting evidence of identity, such as copies of government identification documents; advertisement-customization tools, which allow speakers, for a fee, to target their speech to listeners based upon the listeners' identifying characteristics and behavior; and specific product features that add value to the user experience by curating discourse based upon users' identifying characteristics and behavior. Many other social media companies employ some or all of these practices.

The integration of authenticity regulation into our existing political system has gone virtually unnoticed. For example, when Mark Zuckerberg testified about data privacy to two congressional committees in April 2018, all ninety-eight lawmakers on those committees had personal, verified Facebook Pages.[46] The fact that Zuckerberg was speaking only to lawmakers who had acquiesced in his company's authenticity regulation and were using it for their own benefit calls the lawmakers' independence into question, but few commentators have raised concerns.

Facebook's authenticity regulation continues to evolve in important ways. In the summer of 2017, company executives began publicizing the term "bad actor" to describe individuals and organizations whose speech the company "unpublishes," or bans on the basis of "inauthentic identity."[47] Since then,

---

44 *See* danah boyd, *The Politics of Real Names*, 55 COMMS. ACM 29, 30 (2012).

45 *Terms of Service*, FACEBOOK, https://www.facebook.com/terms.php (last visited Mar. 26, 2020).

46 Robin Opsahl, *Many Lawmakers Questioning Zuckerberg Used Facebook in Their Political Campaigns*, ROLL CALL (Apr. 10, 2018, 5:02 AM), https://www.rollcall.com/news/politics/facebook-advertising-allows-micro-targeted-ads-cambridge-analytica.

47 Mark Zuckerberg, FACEBOOK (June 22, 2017, 1:25 PM), https://www.facebook.com/zuck/posts/10154944663901634 ("[W]e're going to help you remove bad actors and their content quickly . . . ."). Another practice is "shadow banning," which Twitter defines as "deliberately making someone's content undiscoverable to everyone except the person who posted it, unbeknownst to the original poster." Vijaya Gadde & Kayvon Beykpour, *Setting the Record Straight*

company executives have consistently maintained that "bad actors" are responsible for speech-harms on Facebook and are appropriate targets for censorship. The company's executives have referred to "bad actors" as Facebook's "adversaries," and have widely touted their close collaboration with American law enforcement to rout out and silence "bad actors."[48] The term "bad actor" is now used by many companies to villainize those who violate authenticity rules.

In addition, the social media industry has shifted from characterizing inauthentic identity as a *status* to characterizing it as a *behavior*. After the Russian-influence scandal broke, Facebook began expressly prohibiting something it calls "coordinated inauthentic behavior."[49] Over time, other companies, including Twitter, have picked up both the term and the enforcement practices it describes, leading to industry-wide behavioral policing under the authenticity label.

Yet speaker-based regulation presents business risks to the companies that employ it: it puts downward pressure on a key business metric, Monthly Active Users ("MAUs"). Both Facebook and Twitter disclose their MAUs in their securities filings, and investors consider them important to understanding each company's financial performance; high and rising MAUs indicate strong financial prospects. The removal of user accounts for "inauthentic identity" lowers MAUs, depressing this key metric.[50] Starting with its 2016 annual report, and coinciding with a period in which Facebook ramped up its removal of "bad actors," Facebook has steadily increased its estimates of inauthentic accounts as a proportion of MAUs—from approximately 7% of MAUs at the end of 2016 to its most recent estimate of

on Shadow Banning, TWITTER BLOG (July 26, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html. It is not clear whether any social media networks actually employ shadow banning; Twitter has denied doing so. *Id*.

48 *Hearing Before the S. Select Comm. on Intelligence*, 115th Cong. 1–2 (2018) (testimony of Sheryl Sandberg, Chief Operating Officer, Facebook), *available at* https://www.intelligence.senate.gov/sites/default/files/documents/os-ssandberg-090518.pdf [hereinafter Tesitmony of Sheryl Sandberg].

49 *Coordinated Inauthentic Behavior Explained*, FACEBOOK (Dec. 6, 2018), https://about.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior; *see also infra* Part I.B.4.

50 *See* FACEBOOK, 2018 FORM 10-K, at 4, *available at* https://www.sec.gov/Archives/edgar/data/1326801/000132680119000009/fb-12312018x10k.htm (noting that Facebook may take actions "to reduce the number of duplicate or false accounts among our users, which may also reduce our . . . MAU estimates"); Viyaha Gadde, *Confidence in Follower Counts*, TWITTER BLOG (July 11, 2018), https://blog.twitter.com/en_us/topics/company/2018/Confidence-in-Follower-Counts.html ("[S]ome accounts we remove . . . have the potential to impact publicly reported metrics").

16% of MAUs at the end of 2019[51]—revealing a risk that its use of speaker-based strategies could affect its stock price and cost of capital. This conflict makes it potentially useful for Facebook, or any similarly situated company, to employ content-based and speaker-based forms of speech regulation strategically in offsetting ways.

This Part starts by explaining the customization-based business model, and the reliance of the business model, and the model's key metrics, on companies' authenticity rules. Next, it reviews authentic-identity policies and identity verification, which is essentially a system of speech licensing. Finally, it explores the back-office customization system known as micro-targeting, which regulates the flow of information to recipients through micro-targeting, and an assortment of unique product features that companies use to curate discourse by connecting speakers and listeners on the basis of identity.

A. The Business Model: Customization & Analytics

Facebook's revenue model relies almost exclusively on the sale of advertising.[52] Importantly, "advertising" on social media is not limited to traditional advertising—communications designed to market products and services—but includes any speech that receives enhanced distribution for a fee. On Facebook, for example, a user can pay to "boost" posts, increasing the distribution of his or her speech to friends and strangers.[53] When

---

51 *See* FACEBOOK, 2019 FORM 10-K, at 4, *available at* http://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/45290cc0-656d-4a88-a2f3-147c8de86506.pdf (showing that, in the 4th quarter of 2019, "duplicate accounts" represented approximately 11% of worldwide MAUs, and "false accounts" represented approximately 5%).

52 *See* FACEBOOK, 2016 FORM 10-K, at 5, *available at* https://www.sec.gov/Archives/edgar/data/1326801/000132680117000007/fb-12312016x10k.htm ("We generate substantially all of our revenue from selling advertising placements to marketers."); *id.* at 9 ("For 2016, 2015, and 2014, advertising accounted for 97%, 95%, and 92%, respectively, of our revenue."). Interestingly, Mark Zuckerberg has attempted to justify Facebook's advertising-based business model on public interest grounds. In a March 2018 interview with the *New York Times*, Zuckerberg observed that to "bring the world closer together," it is necessary to produce a service "that people can afford." Kevin Roose & Sheera Frenkel, *Mark Zuckerberg's Reckoning: 'This Is a Major Trust Issue'*, N.Y. TIMES (Mar. 21, 2018), https://www.nytimes.com/2018/03/21/technology/mark-zuckerberg-q-and-a.html. "A lot of the people, once you get past the first billion people, can't afford to pay a lot," he explained. *Id.* "Therefore, having it be free and hav[ing] a business model that is ad-supported ends up being really important and aligned [with Facebook's mission]." *Id.*

53 *See* Young Mie Kim et al., *The Stealth Media? Groups and Targets Behind Divisive Issue Campaigns on Facebook*, 35 POL. COMM. 515 (2018) (explaining that "[o]n Facebook . . . a native advertisement appears in News Feeds (as a Sponsored Feed, or Promoted Page) . . . that resembles news, videos,

Facebook and other social media companies talk about "advertising," readers should understand this term to include forms of paid political speech that do not resemble traditional advertising. The social media industry uses the term "organic content" to describe content posted by users that has *not* received any sort of enhanced distribution for a fee. Importantly, Facebook treats paid content (advertising) differently from unpaid (organic) content in ways that make paid expression more impactful.[54]

Facebook's advertisement tools have more or less democratized advertising on social media. Not only are they self-service and extremely easy to use, but they are cheap, allowing users to pay small amounts to communicate nearly anything to a customized audience. Because of its ease of use and low cost, paid political advertising on Facebook is within the means of many citizens. The result is, essentially, an information exchange in which the distribution of information is determined by the amount of money a speaker is willing (or able) to pay, and the identifying data that Facebook can attribute to potential recipients.

1. Front-Office Customization

Facebook utilizes a dual-customization model. One mode of customization is designed to add value to the experience of the retail Facebook user—the recipient of speech. If a recipient demonstrates an interest in something, the platform will deliver more of this type of content to him or her.[55] We might call this "front-office customization," because it is a routine part of the retail user's experience and is apparent to anyone who uses Facebook. Front-office customization is not unique to Facebook. Numerous social media platforms employ front-office customization on the

---

games, memes, or other non-marketing content embedded among regular posts by social media users").

54 As just one example, the company allows users to "snooze" keywords in organic content, but not in paid content. This means that users can designate a word or phrase that they would like filtered out of their feeds—Facebook does this by eliminating from the user's News Feed any post that contains the filtered or "snoozed" term—but Facebook will not filter out paid content using the word. For an explanation of this feature, see Josh Constine, *Facebook Tests 30-Day Keyword Snoozing to Fight Spoilers, Triggers*, TECHCRUNCH (June 27, 2018, 8:00 AM), https://techcrunch.com/2018/06/27/facebook-keyword-snooze/. Facebook has a business reason for treating organic and paid content differently here—it encourages speakers to express themselves through paid content.

55 As Zuckerberg explained in the *Wall Street Journal*, "based on what pages people like, what they click on, and other signals, we create categories—for example, people who like pages about gardening and live in Spain—and then charge advertisers to show ads to that category." Mark Zuckerberg, *The Facts About Facebook*, WALL STREET J. (Jan. 24, 2019), https://www.wsj.com/articles/the-facts-about-facebook-11548374613.

theory that people like it.[56] Because people like to receive customized information, the thinking goes, companies that employ customization will be more likely to attract users, and the platform can turn around and market its large user base to advertisers. In fact, to attract eyeballs, social media companies are in a bit of an arms-race to create and deploy the most user-desired forms of customization.

2. Back-Office Customization

Facebook and other social media companies do not charge users for front-office customization, but they earn vast amounts from "micro-targeting," or what we might call "back-office customization"—customization designed to benefit the paying customers, advertisers. These companies provide paying speakers with complex customization tools that allow them to target communications to only *some* recipients. This is the true heart of social media companies' business models, because it is where advertising dollars are earned.

Back-office customization tools allow paying speakers (advertisers) to target their speech at a subset of Facebook recipients, determined by criteria chosen by the speaker.[57] These tools go well beyond basic demographics, such as gender and "ethnic affinity," to provide extremely granular targeting based on, essentially, whatever the speaker (advertiser) demands.[58] In his

---

56 Interestingly, some evidence suggests this assumption is false. *See, e.g.*, Joseph Turow & Jay Hoofnagle, *Mark Zuckerberg's Delusion of Consumer Consent*, N.Y. TIMES (Jan. 29, 2019), https://www.nytimes.com/2019/01/29/opinion/zuckerberg-facebook-ads.html (summarizing research which found that a "substantial majority" of Americans polled did not want commercial advertisements, news or political advertisements "tailored to your interests").

57 *See* Assurance of Discontinuance at 2, *In re* Facebook, Inc., No. 18-2-18287-5 SEA (Wash. King Cty. Super. Ct. July 24, 2018), *available at* https://agportal-s3bucket.s3.amazonaws.com/uploadedfiles/Another/News/Press_Releases/2018_07_23%20AOD.pdf ("[Facebook] operates a platform that allows advertisers to create and target advertisements using thousands of options based on user interests, including . . . interests in one or more of the following ethnic affinities: 'African American (US)', 'Asian American (US)', 'Hispanic (US-All)', 'Hispanic (US-Bilingual)', 'Hispanic (US-Spanish dominant)', and 'Hispanic (US-English dominant)' . . . ."); Caitlin E. Jokubaitis, *There and Back: Vindicating the Listener's Interests in Targeted Advertising in the Internet Information Economy*, 42 COLUM. J.L. & ARTS 85, 87 (2018) ("Google, Facebook, and LinkedIn[] sustain themselves on a quid pro quo exchange of monetizable user data for a wide array of nominally gratuitous services").

58 *See, e.g.*, Testimony of Colin Stretch, *supra* note 2, at 3 ("Advertisers choose the audience they want to reach based on demographics, interests, behaviors, or contact information."); Balkin, *supra* note 9, at 4. *See generally* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 56–57 (2019) (discussing Facebook's use of "ethnic affinity" in its micro-targeting).

2018 book, *Antisocial Media*, Siva Vaidhyanathan described how, for $200, he used Facebook's micro-targeting tools to promote a post about a podcast:

> I chose to focus the campaign on those who had expressed interest in psychology and neuroscience. I limited the ad placement to those who had an M.D. or a Ph.D. And I excluded those who were younger than thirty years old. This meant I would only reach about three thousand Facebook users. But they would be the right three thousand Facebook users. Just for fun I also excluded any Facebook user who had expressed an interest in the 1970s country music singer Crystal Gayle.[59]

As this reveals, Facebook's customization tools not only allow speakers to create a customized audience for a particular message based upon the recipients' characteristics, but also allow speakers to *exclude* people from that audience based upon their characteristics. In Vaidhyanathan's account, Facebook's advertisement tools allowed him to winnow down his audience to three thousand people and then pay to "serve" his message to only those individuals, for less than seven cents per person.[60]

Facebook's back-office and front-office customization implements what Shoshana Zuboff has called "surveillance capitalism."[61] Zuboff explains that, in surveillance capitalism, human experience is claimed as "free raw material" by companies that transform it into behavioral analytics and predictive technologies.[62] "Much of this new work is accomplished under the banner of 'personalization,'" she writes, a "camouflage" for efforts "that mine the intimate depths of everyday life" for the benefit of companies, like Facebook, that trade in data and data analytics.[63] In surveillance capitalism, individuals' data is commoditized and highly valued, traded among companies, and processed into machine learning, which can be sold.

Facebook's customization is made possible in part by its authenticity rules. Its back-office customization tools are effective, and generate high profits, because they leverage information gleaned about Facebook users who must use only their "authentic" identities and provide only accurate

---

59 SIVA VAIDHYANATHAN, ANTISOCIAL MEDIA 88 (2018).

60 In a different part of his book, Vaidhyanathan describes himself as the target of back-office customization, rather than its initiator. After he disclosed his "married" status on Facebook, Vaidhyanathan writes, the "advertising spaces on my Facebook page filled up with advertisements for services that invited me to contact women for the purpose of having an affair." *Id.* at 56.

61 ZUBOFF, *supra* note 10, at 8–9; *see also* Balkin, *supra* note 9, at 2 ("Data collection and analysis allow targeted advertising, which allows more efficient advertising campaigns, which allow greater revenues.").

62 ZUBOFF, *supra* note 10, at 8–9.

63 *Id.* at 53 ("Under this new regime, the precise moment at which our needs are met [through customization] is also the precise moment at which our lives are plundered for behavioral data, and all for the sake of others' gain.").

information about themselves. The scope of the information that Facebook collects on each user—and perhaps even on non-users—is significant.[64] Moreover, the reach of advertising on Facebook is measurable because the company's authenticity policies ensure that every person has only one account; as a result, the company can say with accuracy how many individuals receive the advertising it transmits. Company executives expressly link the effectiveness of customization to Facebook's "real identity" approach and have described the "real identity" policy to financial analysts as a "significant advantage" that Facebook enjoys over its competitors.[65]

While Facebook's authenticity regulation is the substrate upon which its business model is built, its commercial potential extends beyond the sale of advertising. Technology-industry observers have long contended that Facebook sees identity itself as a business opportunity.[66] Since going public in 2012, Facebook has acquired several companies that specialize in biometric identity verification technology, suggesting that the company is at least leaving open the possibility that it will move into this space.[67] In early 2017, the company launched a feature that provides encrypted recovery tokens for other websites—"a way for Facebook to convince users to center

64 *See, e.g.*, *id.* at 252 (describing Facebook's strides in biometrics).

65 Conference Call of Facebook Executives on First Quarter 2017 Earnings 24 (May 3, 2017), *available at* https://s21.q4cdn.com/399680738/files/doc_financials/2017/Q1-'17-Earnings-transcript.pdf (recording Sheryl Sandberg statement that: "We think that targeting and measurement are significant competitive advantages for us. . . . We believe that because people are sharing interests because people are themselves their real identity on the Facebook platform, we have a significant advantage"). In a 2012 internal Facebook email, Mark Zuckerberg made explicit the connection between Facebook's business prospects and its ability to coax users to share identifying information: "Sometimes the best way to enable people to share something is to have a developer build a special purpose app or network for that type of content and to make that app social by having Facebook plug into it. That may be good for the world, but it's not good for us unless people also share back to Facebook and that content increases the value of our network." *DealBook Briefing: Inside the Emails Facebook Never Thought You'd Read*, N.Y. TIMES (Dec. 6, 2018), https://www.nytimes.com/2018/12/06/business/dealbook/facebook-email-data.html.

66 *See, e.g.*, Donald Melanson, *This is Your Life: Facebook and the Business of Identity*, ENGADGET (July 16, 2013), https://www.engadget.com/2013/07/16/facebook-and-the-business-of-identity (referring to Facebook as "your single sign-on internet identity"); Semil Shah, *Another Reason Facebook Wants a Web of Real Identities: Commerce*, TECHCRUNCH (Aug. 1, 2011, 4:22 PM), https://techcrunch.com/2011/08/01/facebook-real-identities-commerce/ ("for networks like Facebook, the game is to encourage its users to leverage their real identities online so that Facebook can accelerate its ability to *power online transactions*").

67 *See* ZUBOFF, *supra* note 10, at 129, 242–54 (discussing "behavioral surplus capture" in "the real world" by companies like Google and Facebook).

their online identity around their Facebook profile," as one writer put it.[68] A year later, only months before announcing its new identity-verification requirements for political speech, Facebook bought start-up Confirm.io, known for implementing biometric screening systems. An article about the acquisition in the trade publication *TechCrunch* suggested that Facebook's ambition is to "serve as your ID card in some situations," a service for which it could likely charge a fee.[69] In 2019, Facebook announced its participation in the Libra Association, which will launch a cryptocurrency, and its development of a new product, the Calibra wallet, for utilizing the Libra cryptocurrency.[70] These lines of business will capitalize on the company's ability to securely tie cyber cash flows to online identities.

B. Authentic Identity Rules

In their rules, social media companies typically distinguish between "authentic" and "inauthentic" users and restrict the flows of speech to and from "inauthentic" users.[71] A speaker who employs a false name or identity, or provides the company with false or misleading information about him or herself, is an "inauthentic" speaker.[72] Twitter prohibited its users from

---

68 Kate Conger, *Facebook Challenges Email for Control of Your Online Identity*, TECHCRUNCH (Jan. 30, 2017, 12:50 PM), https://techcrunch.com/2017/01/30/facebook-challenges-email-for-control-of-your-online-identity/.

69 Josh Constine, *Facebook Acquires Biometric ID Verification Startup Confirm.io*, TECHCRUNCH, (Jan. 23, 2018, 4:36 PM), https://techcrunch.com/2018/01/23/facebook-confirm-io/. The same journalist, a longtime technology-industry reporter with a focus on Facebook, recently observed that "Facebook has become the identity layer for the internet." Josh Constine, *'Facebook Avatars' is Its New Clone of Snapchat's Bitmoji*, TECHCRUNCH (May 7, 2018, 8:20 PM), https://techcrunch.com/2018/05/07/facebook-avatars/.

70 *See Coming in 2020: Calibra*, FACEBOOK NEWSROOM (June 18, 2020), https://newsroom.fb.com/news/2019/06/coming-in-2020-calibra/; *Libra White Paper*, LIBRA, https://libra.org/en-US/white-paper/ (last visited Mar. 26, 2020); Josh Constine, *Facebook Announces Libra Cryptocurrency: All You Need to Know*, TECHCRUNCH (June 18, 2019, 5:01 AM), https://techcrunch.com/2019/06/18/facebook-libra/.

71 *See, e.g.*, Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/ (noting that Facebook has "worked hard to establish authenticity as a social norm"); *Instagram Community Guidelines,* INSTAGRAM HELP CTR., https://help.instagram.com/477434105621119 (stating that users must "[r]emember to post authentic content") (last visited Jan. 4, 2020). Twitter streamlined its "Twitter Rules" in June 2019; the new rules include a section entitled "Authenticity," with subheadings on "Platform Manipulation and Spam," "Election Integrity," "Impersonation," and "Copyright and Trademark." *See Twitter Rules*, TWITTER HELP CTR., https://help.twitter.com/en/rules-and-policies/twitter-rules (last visited Jan. 16, 2020).

72 As explained more fully below, a user who engages in "inauthentic *behavior*" can also find him or herself subject to speech regulation, as Facebook has been expanding its identity-based rules to incorporate behavior as a facet of identity. *See infra* Part I.B.4.

registering or creating "fake and misleading accounts" but did not specifically use the word "authenticity" in its "Twitter Rules" until June 2019.[73] In general, Twitter is more permissive than Facebook about the use of pseudonyms and multiple accounts.[74] Like other companies, when Facebook determines that a speaker is inauthentic, it may shut down the speaker's account, remove the speaker's speech, and prevent the speaker from engaging in future speech activity on its network.[75] It may also transmit its finding of "inauthentic" identity to other social media companies, which in turn censor that speaker.[76] Once a user is removed from a platform for inauthenticity, that user generally may not return to the platform.[77]

Facebook's policies and rules about "authentic" identity are distilled in its terms of service, its Community Standards, and in other documents and publications. Its terms of service lay out the basic contours.[78]

---

73 *See* Del Harvey, *Making Twitter's Rules Easier to Understand*, TWITTER BLOG (June 7, 2019), https://blog.twitter.com/en_us/topics/company/2019/rules-refresh.html; *Twitter Rules*, TWITTER HELP CTR., https://help.twitter.com/en/rules-and-policies/twitter-rules (last visited Jan. 16, 2020). After the revisions, a hyperlink found under the heading "Authenticity," labeled "Platform manipulation and spam" brings up Twitter's September 2019 "Platform Manipulation and Spam Policy," which prohibits "inauthentic engagements" and "coordinated activity," and states: "You can't mislead others on Twitter by operating fake accounts." *Platform Manipulation and Spam Policy*, TWITTER HELP CTR., https://help.twitter.com/en/rules-and-policies/platform-manipulation (last visited Jan. 5, 2020).

74 *See Parody, Newsfeed, Commentary and Fan Account Policy*, TWITTER HELP CTR., https://help.twitter.com/en/rules-and-policies/parody-account-policy (last visited Jan. 16, 2020).

75 *See, e.g.*, ALEX SCHULTZ & GUY ROSEN, FACEBOOK, UNDERSTANDING THE FACEBOOK COMMUNITY STANDARDS ENFORCEMENT REPORT 22 (2018), https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf ("When we identify a fake account, we disable it so it's no longer visible and its owner can't log in.").

76 *See, e.g.*, Testimony of Colin Stretch, *supra* note 2, at 7 ("[Facebook is] reaching out to leaders in our industry . . . to share information on bad actors . . . [to] make sure they stay off all platforms"); Nathaniel Gleicher, *How We Work With Our Partners to Combat Information Operations*, FACEBOOK NEWSROOM (Nov. 13, 2018), https://newsroom.fb.com/news/2018/11/last-weeks-takedowns/ ("[W]e've worked closely with our fellow tech companies . . . to deal with the threats . . . ."); Tony Romm & Craig Timberg, *Facebook Suspends 'Inauthentic' Iranian Accounts that Criticized Trump and Spread Divisive Political Messages*, WASH. POST (Oct. 26, 2018, 2:24 PM), https://www.washingtonpost.com/technology/2018/10/26/facebook-suspends-inauthentic-iranian-accounts-that-criticized-trump-spread-divisive-political-messages/ ("Twitter said it had removed a small number of accounts based on information Facebook supplied.").

77 *See, e.g.*, *Platform Manipulation and Spam Policy*, TWITTER HELP CTR., https://help.twitter.com/en/rules-and-policies/platform-manipulation (noting that Twitter considers it a "severe violation" of its rules when a user "creat[es] accounts to replace or mimic suspended accounts").

78 *See Terms of Service*, FACEBOOK, https://www.facebook.com/terms (last visited Mar. 26, 2020); *see also What Names Are Allowed on Facebook?*, FACEBOOK HELP CTR., https://www.facebook.com/help/112146705538576 (last visited Jan. 4, 2020). For examples involving other social media companies, see, e.g., *Snap Inc. Terms of Service*, SNAP INC., https://www.snap.com/en-US/terms/ (last visited Mar. 26, 2020) (limiting users to a single account); *Community Guidelines*, SNAP

It is Facebook's longstanding practice to purge its network of "inauthentic" users.[79] More than a decade ago, law professor James Grimmelman observed that Facebook applied its prohibition against false identity "rigorously, almost to the point of absurdity."[80] Facebook allows users to report others for employing fake names, which by late 2014 had led to "several hundred thousand fake name reports" submitted to the company weekly.[81]

Facebook's authenticity policies have been criticized for discriminating against certain kinds of identities, including Native Americans whose names do not conform to popular conventions,[82] and transgender individuals who have been accused of employing identity-based deception.[83]

The rules have also been criticized by domestic violence survivors and members of the LGBTQ community for potentially exposing people to harm.[84] In October 2015, a coalition of civil society organizations wrote an open letter to Facebook asking it to let users employ pseudonyms in situations where "using an every day [sic] name would put a user in danger" and to

---

INC., https://www.snap.com/en-US/community-guidelines ("Don't pretend to be someone you're not—this includes your friends, celebrities, brands, or other organizations—or attempt to deceive people about who you are.") (last visited Mar. 26, 2020); *Policy on Impersonation*, YOUTUBE HELP, https://support.google.com/youtube/answer/2801947 (last visited Mar. 26, 2020).

79 *See, e.g.*, Barbara Ortutay, *A Facebook Identity Crisis: A Familiar Face, But the Name?*, CHARLESTON GAZETTE, May 19, 2009, at P6A (recounting the story of a woman whose profile Facebook erroneously removed as fake and noting that "[t]o make sure people can't set up accounts with fake names, the site has a long, constantly updated 'blacklist' of names that people can't use," including names "that sound fake, like Batman, or names tied to current events"); *see also infra* notes 94–96 (describing purges of inauthentic accounts from 2016 to present).

80 James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1143 (2009).

81 Chris Cox, FACEBOOK (Oct. 1, 2014, 2:22 PM), https://www.facebook.com/chris.cox/posts/10101301777354543.

82 *See, e.g.*, Abby Phillip, *Online "Authenticity" and How Facebook's 'Real Name' Policy Hurts Native Americans*, WASH. POST (Feb. 10, 2015), https://www.washingtonpost.com/news/morning-mix/wp/2015/02/10-online-authenticity-and-how-facebooks-real-name-policy-hurts-native-americans/.

83 *See, e.g.*, Brittney McNamara, *This Person Says Facebook's 'Authentic Name' Policy is 'Anti-Trans,'* TEEN VOGUE (Nov. 7, 2017), https://www.teenvogue.com/story/facebook-authentic-name-policy (describing the Facebook experience of a non-binary member of the clergy of the United Methodist Church).

84 *See* Samantha Allen, *How Facebook Exposes Domestic Violence Survivors*, DAILY BEAST (May 20, 2015), https://www.thedailybeast.com/how-facebook-exposes-domestic-violence-survivors (recounting the story of a domestic violence survivor whose account Facebook shut down for using a pseudonym; after she complied with Facebook's requirement to reopen the account under her legal name, her abusive ex-husband found her and began harassing her almost immediately, despite having been out of touch for 18 years); Reed Albergotti, *Facebook Versus the Drag Queens*, WALL STREET J. (Sept. 12, 2014, 8:44 PM), https://blogs.wsj.com/digits/2014/09/12/facebook-versus-the-sisters-of-perpetual-indulgence/ ("Recently, [Facebook] took aim at performers who use stage names instead of legal names in their Facebook profiles, forcing them to use their real identities.").

allow users to confirm their identities "without submitting government ID."[85] While both groups were concerned that Facebook's real-name rules could put them at risk, the LGBTQ community also argued that its members had a right to define their own identities. Facebook responded by announcing, at the end of 2015, that it would begin testing small exceptions to its authenticity requirements, but it essentially left its rules intact.[86]

More recently, it has become apparent that authenticity requirements can put pro-democracy and human rights activists at risk.[87] Facebook has suggested that it makes case-by-case exceptions for activists.[88]

1. Inauthenticity as a Business Risk

Social media companies have tended to treat inauthentic user identity as a business risk requiring enforcement efforts. Facebook has consistently linked authenticity to two of its major business risks: its ability to maintain its brands, and the accuracy of its user metrics. The company's securities filings since 2012 have warned that Facebook's brands may be "negatively affected" by "users acting under false or inauthentic identities," presumably because

85 *Open Letter to Facebook About its Real Names Policy*, ELECTRONIC FRONTIER FOUND. (Oct. 5, 2015), https://www.eff.org/document/open-letter-facebook-about-its-real-names-policy.

86 *See* Justin Osofsky & Todd Gage, *Community Support FYI: Improving the Names Process on Facebook*, FACEBOOK NEWSROOM (Dec. 15, 2015), https://newsroom.fb.com/news/2015/12/community-support-fyi-improving-the-names-process-on-facebook/ (distinguishing that changes were being "tested on a limited basis in the US only"); Eva Galperin & Wafa Ben Hassine, *Changes to Facebook's 'Real Names' Policy Still Don't Fix the Problem*, ELECTRONIC FRONTIER FOUND. (Dec. 18, 2015), https://www.eff.org/deeplinks/2015/12/changes-facebooks-real-names-policy-still-dont-fix-problem (describing Facebook's response as "rearranging chairs on the Titanic").

87 *See, e.g.*, Chloe Tennant, *Russia Charges Activist for a Facebook Post*, HUM. RTS. WATCH (Mar. 22, 2019, 1:39 PM), https://www.hrw.org/news/2019/03/22/russia-charges-activist-facebook-post (describing how Russian prosecutors filed charges against an activist who posted an infographic comparing the prices of items in Russia in 2009 and 2019); Emily Price, *Twitter and Human Rights: A Complicated Story*, HUM. RTS. FIRST (Mar. 26, 2014) https://www.humanrightsfirst.org/blog/twitter-and-human-rights-complicated-story (recounting how a Kuwaiti court sentenced Mohammed Eid al-Ajmi to five years in prison for criticizing Kuwait's Amir in a tweet). Open Democracy reported that Vietnamese activists were denied the use of pseudonyms by Facebook, which demanded evidence of their identities and then changed their account names to match their legal identification without notifying the activists that it would do so. *See* Brett Solomon, *What Can Social Media Platforms Do For Human Rights?*, OPENDEMOCRACY (Oct. 26, 2015) https://www.opendemocracy.net/en/what-can-social-media-platforms-do-for-human-rights/ (noting that several Vietnamese writers and activists were not allowed to use their pen names on Facebook).

88 *See, e.g.*, Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/ (describing Facebook's efforts to remove Pages controlled by the IRA surrounding the 2016 presidential election).

such users could cause the company reputational harm.[89] Facebook's financial reports have also warned investors that "real or perceived inaccuracies" in its "user and other metrics" may harm its business, and have disclosed information about the numbers of "duplicate accounts" and "false accounts" that Facebook believes exist.[90] The company's user metrics shed important light on its engagement and growth and are closely followed by investors; significant changes in its user metrics have caused stock analysts to raise concerns about the company's prospects. Facebook's disclosures of business risks related to inauthenticity suggest that Facebook's business focus on the issue has been serious and consistent since its earliest days as a public company.

2. The 2016 Election and its Aftermath

In 2017, the American public learned about a significant foreign-state-sponsored campaign to influence 2016 federal elections, waged on social media networks including Facebook, Instagram, and Twitter.[91] In connection with this news, these companies purged the Russian-linked accounts; Facebook stated that it had shut down the Russian-linked accounts

89 *See, e.g.*, FACEBOOK, 2018 FORM 10-K, at 14; FACEBOOK, 2017 FORM 10-K, at 13; FACEBOOK, 2016 FORM 10-K, at 12; FACEBOOK, 2015 FORM 10-K, at 12; FACEBOOK, 2014 FORM 10-K, at 15; FACEBOOK, 2013 FORM 10-K, at 18; FACEBOOK, 2012 FORM 10-K, at 21; *Annual Reports*, FACEBOOK INVESTOR RELATIONS, https://investor.fb.com/financials/default.aspx (follow hyperlinks to corresponding Form 10-K for each fiscal year) (last visited Mar. 26, 2020).

90 Facebook further divides "false accounts" into two subsets: "user-misclassified accounts" and "undesirable accounts." *See supra* note 89 (defining the differences between the subsets of false accounts in annual reports). In Facebook's lexicon, a "duplicate account" is an account "that a user maintains in addition to his or her principal account." FACEBOOK, 2016 FORM 10-K, at 4, *available at* https://www.sec.gov/Archives/edgar/data/1326801/000132680117000007/fb-12312016x10k.htm A "user-misclassified account" is "where users have created personal profiles for a business, organization, or non-human entity such as a pet." *Id.* An "undesirable account" is a user profile "that we determine [is] intended to be used for purposes that violate our terms of service, such as spamming." *Id.*

91 For details about the interference campaign, see generally, ROBERT S. MUELLER, III, U.S. DEP'T OF JUSTICE, 1 REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION (2019) [hereinafter The Mueller Report]. Although Facebook did not publicly disclose what it knew about Russian-linked accounts on its network until September 2017, we know today that Facebook was aware, before the 2016 election, of specific cyber threats from "actors with ties to Russia." *See* Testimony of Sheryl Sandberg, *supra* note 48, at 2. Sheryl Sandberg told the Senate Select Committee on Intelligence in September 2018 that Facebook had "detected and mitigated several threats from actors with ties to Russia" before Election Day in 2016. *Id.* She also stated that Facebook "saw some new behavior—namely, the creation of fake personas that were then used to seed stolen information to journalists." *Id.*

and Pages because "[w]e don't allow inauthentic accounts on Facebook."[92] Over subsequent weeks, other Facebook executives reiterated this point; Sandberg told an interviewer on camera that Facebook would have permitted most of the offending Russian advertisements "if they were run by legitimate people."[93] These statements underscored the company's choice to emphasize authenticity harms, rather than substantive harms, when it discussed the IRA's electoral interference.

Soon afterward, Facebook began a series of well-publicized purges of fake accounts.[94] Twitter has also made public its purges of fake accounts.[95] The

---

92 Alex Stamos, *An Update on Information Operations at Facebook*, FACEBOOK NEWSROOM (Sept. 6, 2017), https://newsroom.fb.com/news/2017/09/information-operations-update/.

93 *Exclusive Interview with Facebook's Sheryl Sandberg*, AXIOS (Oct. 12, 2017), https://www.axios.com/exclusive-interview-with-facebooks-sheryl-sandberg-1513306121-64e900b7-55da-4087-afee-92713cbbfa81.html. Sandberg was essentially reiterating points that were made days earlier in a Facebook blog post:

> We require authenticity regardless of location. If Americans conducted a coordinated, inauthentic operation—as the Russian organization did in this case—we would take their ads down, too. However, many of these ads did not violate our content policies. That means that for most of them, if they had been run by authentic individuals, anywhere, they could have remained on the platform.

Elliot Schrage, *Hard Questions: Russian Ads Delivered to Congress*, FACEBOOK NEWSROOM (Oct. 2, 2017), https://newsroom.fb.com/news/2017/10/hard-questions-russian-ads-delivered-to-congress/.

94 In April 2018, Facebook removed "more than 270" Pages and accounts "controlled by the IRA." Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/ (highlighting the 70 Facebook accounts, 138 Facebook pages, and 65 Instagram accounts removed for lack of authenticity); Testimony of Sheryl Sandberg, *supra* note 48, at 2. A few months later, in July, Facebook revealed that it had again detected "Bad Actors" on Facebook and Instagram. *Removing Bad Actors on Facebook*, FACEBOOK NEWSROOM (July 31, 2018), https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/; *see also* Testimony of Sheryl Sandberg, *supra* note 48, at 3 (noting that "whoever set up these accounts went to greater lengths to obscure their true identities than the IRA did in 2016"). As part of this purge, the company removed eight Facebook Pages and seventeen Facebook profiles that it said violated its "ban on coordinated inauthentic behavior." Nathaniel Gleicher, *What We've Found So Far*, FACEBOOK NEWSROOM (July 31, 2019), https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/#what-weve-found. In August, Facebook again announced that it had detected an influence campaign and had removed content. *Taking Down More Coordinated Inauthentic Behavior*, FACEBOOK NEWSROOM (Aug. 21, 2018), https://newsroom.fb.com/news/2018/08/more-coordinated-inauthentic-behavior/. The content in question was found on Facebook and Instagram. *Id.* Facebook also announced that it had removed an unspecified number of Pages and accounts that "can be linked to sources that the U.S. government has previously identified as Russian military intelligence services." Testimony of Sheryl Sandberg, *supra* note 48, at 4.

95 *See* Yoel Roth, *Information Operations on Twitter: Principles, Process, and Disclosure*, TWITTER BLOG (June 13, 2019), https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html (describing Twitter's efforts to remove state-backed accounts).

purges continue to this day and are commonly disclosed to the public by each company in a press release.[96]

In 2018, Facebook began publishing a "Community Standards Enforcement Report" that included details on its removal of fake accounts and spam.[97] According to the November 2018 Report, Facebook removed 694 million fake accounts in the fourth quarter of 2017, 583 million fake accounts in the first quarter of 2018, 800 million fake accounts in the second quarter of 2018, and 754 million fake accounts in the third quarter of 2018.[98] It's not clear if these numbers reflect the growth of fake accounts on Facebook, or instead improvements by Facebook in its ability to identify and remove fake accounts.

What is clear, however, is that by the end of 2018, Facebook was removing more speech for authenticity violations than for content violations. In the third quarter of 2019, for example, Facebook "took action on" 91.7 million items for all content-related violations combined, including "bullying and harassment," "violence and graphic content," and hate speech.[99] However, in the same quarter, it removed 1.7 billion fake accounts.[100] What

---

96 As of this writing, the most recent purge of fake accounts by Facebook was reported on March 12, 2020. *See* Nathaniel Gleicher, *Removing Coordinated Inauthentic Behavior*, FACEBOOK NEWSROOM (Mar. 12, 2020), https://about.fb.com/news/2020/03/removing-coordinated-inauthentic-behavior-from-russia/. The most recent purge of fake accounts by Twitter was reported on August 19, 2019. *See Information Operations Directed at Hong Kong*, TWITTER BLOG (Aug. 19, 2019), https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html (recounting Twitter's efforts to suspend accounts originating from mainland China that aimed to sow political discord in Hong Kong).

97 A preliminary report was published in May 2018, and a more detailed report in November. *See* FACEBOOK NEWSROOM, DATA SNAPSHOT: FACEBOOK'S COMMUNITY STANDARDS ENFORCEMENT REPORT (2018), https://fbnewsroomus.files.wordpress.com/2018/11/cser-data-snapshot-nov2018-1.jpg (showing that Facebook has increasingly taken action against spam posts as of November 2018) [hereinafter DATA SNAPSHOT].

98 *Id.*

99 *See Community Standards Enforcement Report*, FACEBOOK, https://transparency.facebook.com/community-standards-enforcement (last visited Mar. 31, 2020); *see also* DATA SNAPSHOT, *supra* note 97 (showing that, in the "How much content did Facebook take action on?" column, in the third quarter of 2018, Facebook "took action on" 62.9 million items for content-related violations, such as "adult nudity and sexual activity," "violence and graphic content," and hate speech, and 1.985 billion items in the categories "fake accounts" and "spam"). Unfortunately, the Report provided no information on content removed for "coordinated inauthentic behavior."

100 *Community Standards Enforcement Report*, FACEBOOK, https://transparency.facebook.com/community-standards-enforcement#fake-accounts (last visited Mar. 31, 2020). In fact, the Report's figures on fake accounts and spam appear to reflect takedowns of the whole account, though the Report uses the word "item" to describe all types of removed speech. This could mean that when disclosing content-based takedowns, the Report treated a single post containing bad content as an "item," but when disclosing identity-based takedowns, it treated a whole fake account as a single "item." In

is more, the report provided figures only for *some* identity-based prohibitions—fake accounts and spam—and revealed nothing about other types of speech that Facebook has said it removes for authenticity violations, such as "coordinated inauthentic behavior."

3. Bad Actors and Bad-Faith Actors

In the summer of 2017, at a time when Facebook was learning internally about the actions of Russian-linked groups,[101] company executives began popularizing the term "bad actor" to describe individuals and organizations whose speech the company "unpublishes" or bans on the basis of "inauthentic identity."[102] Since then, company executives consistently have maintained that "bad actors" are responsible for harms caused by speech on Facebook and are appropriate targets for censorship.[103] In public remarks, both of Facebook's top executives, Zuckerberg and Sandberg, have referred to "bad actors" as Facebook's "adversaries."[104]

---

addition, the Report made clear that it did not include "blocked attempts" to create fake accounts in its reported figures. *Id.*

101 The *New York Times* has recounted how Facebook's security head, Alex Stamos, informed the company's Audit Committee in September 2017 about internal findings on the role of Russian-linked groups on the company's network. *See* Sheera Frenkel et al., *Delay, Deny and Deflect: How Facebook's Leaders Leaned Out in Crisis*, N.Y. TIMES (Nov. 14, 2018), https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html (recounting the fallout from Alex Stamos' discoveries within Facebook).

102 The earliest use of "bad actor" by a Facebook executive appears to be from October 2014, when Chris Cox apologized to the LGBTQ community in a blog post for the company's real-name policy. Cox wrote that "99 percent" of the "several hundred thousand fake name reports" that Facebook processed "every single week" "are bad actors doing bad things: impersonation, bullying, trolling, domestic violence, scams, hate speech, and more." Chris Cox, FACEBOOK (Oct. 1, 2014, 2:22 PM), https://www.facebook.com/chris.cox/posts/10101301777354543. The use of the term by Facebook and its executives became much more common starting in June of 2017. *See, e.g.*, Conference Call of Facebook Executives on Second Quarter 2017 Earnings 2 (July 26, 2017), *available at* https://s21.q4cdn.com/399680738/files/doc_financials/2017/Q2/Q2-'17-Earnings-call-transcript.pdf (documenting the comments of Mark Zuckerberg); Conference Call of Facebook Executives on Third Quarter 2017 Earnings 2–3 (Nov. 1, 2017), *available at* https://s21.q4cdn.com/399680738/files/doc_financials/2017/Q3/Q3-'17-Earnings-call-transcript.pdf (noting Zuckerberg's use of "bad actors" three times); *Exclusive Interview with Facebook's Sheryl Sandberg*, AXIOS (Oct. 12, 2017), https://www.axios.com/exclusive-interview-with-facebooks-sheryl-sandberg-1513306121-64e900b7-55da-4087-afee-92713cbbfa81.html (recording a statement by Sheryl Sandberg in which she used the term).

103 *See, e.g.*, Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/ (describing the Russia-based Internet Research Agency as a "bad actor").

104 Testimony of Sheryl Sandberg, *supra* note 48, at 1–2; Alex Stamos, *How Much Can Companies Know About Who's Behind Cyber Threats?*, FACEBOOK NEWSROOM (July 31, 2018), https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/#whos-behind-

Executives at other social media companies quickly adopted the use of "bad actors,"[105] or sometimes "bad-faith actors."[106] The term moved to the news media,[107] and the presidential administration,[108] and is now commonly used by speakers outside of the industry.

---

cyber-threats ("[Facebook] face[s] determined, well-funded adversaries who will never give up and are constantly changing tactics. It's an arms race and we need to constantly improve too. It's why we're investing heavily in more people and better technology to prevent bad actors misusing Facebook—as well as working much more closely with law enforcement and other tech companies to better understand the threats we face."); Nathaniel Gleicher, *Removing Bad Actors on Facebook*, FACEBOOK NEWSROOM (July 31, 2018), https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/.

105 *See, e.g.*, Ronan Costello, *Working Together for a Safer Internet*, TWITTER BLOG (Feb. 5, 2019), https://blog.twitter.com/en_us/topics/company/2019/SaferInternetDay2019.html ("[Twitter will] continue to make it more difficult for bad actors to create spammy or fake accounts that manipulate our platform"); Donie O'Sullivan, Drew Griffin & Curt Devine, *In Attempt to Sow Fear, Russian Trolls Paid for Self-Defense Classes for African Americans*, CNN BUSINESS (Oct. 18, 2017, 9:30 PM), http://money.cnn.com/2017/10/18/media/black-fist-russia-self-defense-classes/index.html (demonstrating an instance of the technology company Eventbrite using "bad actor"); *Expanding Our Work Against Abuse of Our Platform*, YOUTUBE OFFICIAL BLOG (Dec. 4, 2017), https://youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html (using the term "bad actor" three times); *see also Faster Removals and Tackling Comments—An Update on What We're Doing to Enforce YouTube's Community Guidelines*, YOUTUBE OFFICIAL BLOG (Dec. 13, 2018), https://youtube.googleblog.com/2018/12/faster-removals-and-tackling-comments.html ("The vast majority of attempted abuse [on YouTube] comes from bad actors . . . .").

106 *See, e.g.*, Vinja Gadde & Kayvon Beykpour, *Setting the Record Straight on Shadow Banning*, TWITTER BLOG (July 26, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html (noting that Twitter must "address bad-faith actors who intend to manipulate or detract from healthy conversation").

107 Major media outlets that have integrated the term "bad actors" into their news reporting include the *New York Times*, the *Wall Street Journal*, and the *Washington Post*. *See, e.g.*, Brian X. Chen, *Unknown Tech Brands Aren't Like Groceries. Don't Just Grab Them.*, N.Y. TIMES (Apr. 5, 2018), https://www.nytimes.com/2018/04/04/technology/personaltech/beware-unknown-tech-brands.html ("Smartphones . . . are embedded with microphones, motion sensors and cameras that can spy on your every move if corrupted by a bad actor"); Dustin Volz, *U.S. Girds for Possible Russian Meddling on Election Day*, WALL STREET J. (Nov. 5, 2018, 7:30 AM), https://www.wsj.com/articles/u-s-girds-for-possible-russian-meddling-on-election-day-1541421000 (describing how one model of ballot-counting machine "has a flaw detected over a decade ago that could give a bad actor with physical access the ability to change a vote tally"); Lori Aratani, *U.S. Customs Officials Thwart Egyptian Locust Invasion*, WASH. POST (Dec. 4, 2018, 4:55 PM), https://www.washingtonpost.com/transportation/2018/12/04/us-customs-officials-thwart-egyptian-locust-invasion/ ("Like many other types of locusts, the Egyptian tree locust is a bad actor, a very bad actor.").

108 Sarah Huckabee Sanders, then the White House Press Secretary, attached the "bad actor" label to a former FBI official, Andrew McCabe. *See* Ben Yagoda, *'Bad Actor' Is Everywhere. When Did It Start?*, CHRON. HIGHER EDUC. (Mar. 26, 2018), https://www.chronicle.com/blogs/linguafranca/2018/03/26/bad-actor-is-everywhere-when-did-it-start/; *see also Excerpts from Trump's Interview with the Times*, N.Y. TIMES (Feb. 1, 2019), https://www.nytimes.com/2019/02/01/us/politics/trump-interview-transcripts.html (quoting President Trump as stating that "Iran is a bad actor"); Donald J. Trump, Pres. of the United States, Remarks on Combatting Drug Demand and the Opioid Crisis

4. Coordinated Inauthentic Behavior

Over time, social media companies have made a shift from characterizing inauthentic identity as a status to characterizing it as a behavior. For example, after the 2016 election, Facebook began prohibiting "coordinated inauthentic behavior," which, according to Sheryl Sandberg, is "when multiple accounts—including both fake and authentic accounts—work together to mislead people."[109] Facebook's current Community Standards tell users to not:

- Engage in or claim to engage in Inauthentic Behavior, which is defined as the use of Facebook or Instagram assets (accounts, pages, groups, or events), to mislead people or Facebook:
  - about the identity, purpose, or origin of the entity that they represent
  - about the popularity of Facebook or Instagram content or assets
  - about the purpose of an audience or community
  - about the source or origin of content
  - to evade enforcement under our Community Standards
- Engage in, or claim to engage in Coordinated Inauthentic Behavior, defined as the use of multiple Facebook or Instagram assets, working in concert to engage in Inauthentic Behavior (as defined above), where the use of fake accounts is central to the operation
- Engage in or claim to engage in Foreign or Government Interference, which is Coordinated Inauthentic Behavior conducted on behalf of a foreign or government actor[110]

Thus, under Facebook's evolving rules, it is now possible to have an authentic identity on Facebook but to nonetheless violate the company's rules against inauthentic *behavior*. Individuals who engage in coordinated

(Oct. 26, 2017), *available at* https://www.whitehouse.gov/briefings-statements/remarks-president-trump-combatting-drug-demand-opioid-crisis/ (noting that the President was looking at bringing "major lawsuits" against "bad actors" involved in the opioid trade).

109 Testimony of Sheryl Sandberg, *supra* note 48, at 3; *see also* Sacha Pfeiffer, *Inside Saudi Arabia's Disinformation Campaign*, NPR (Aug. 10, 2019, 8:34 AM), https://www.npr.org/2019/08/10/750086287/inside-saudi-arabias-disinformation-campaign (explaining that coordinated inauthentic behavior is "Facebook's catch-all term for groups of accounts that work together to mislead about either who they are or what they're doing"). Sandberg stated that coordinated inauthentic behavior "is not allowed because we don't want organizations or individuals creating networks of accounts that misinform people about who they are or what they're doing." Testimony of Sheryl Sandberg, *supra* note 48, at 3.

110 *Facebook Community Standards: Inauthentic Behavior*, FACEBOOK https://www.facebook.com/communitystandards/inauthentic_behavior/ (last visited Jan. 4, 2020). These standards were updated to "explicitly ban coordinated inauthentic behavior" in late 2018. *See* Telephone Interview of Facebook Officials (Nov. 15, 2018), *available at* https://fbnewsroomus.files.wordpress.com/2018/11/call-transcript-11_15_2018.pdf.

inauthentic behavior are "bad actors" and are subject to censorship on that basis.[111]

Twitter also picked up the use of the term "coordinated inauthentic behavior"[112] and has said that it considers behavior a "signal" for "serving healthy public conversation."[113] The company has explained that it uses the following signals to identify "bad-faith actors":

- Specific account properties that indicate authenticity (e.g. whether you have a confirmed email address, how recently your account was created, whether you uploaded a profile image, etc)
- What actions you take on Twitter (e.g. who you follow, who you retweet, etc)
- How other accounts interact with you (e.g. who mutes you, who follows you, who retweets you, who blocks you, etc)[114]

The second and third signals are based on expressive behavior, although the third signal focuses not on a speaker's own expressive behavior, but the expressive behavior of others.[115]

In 2019, Twitter publicly apologized for purging accounts belonging to pro-democracy activists tweeting about China in advance of the thirtieth anniversary of the Tiananmen Square uprising.[116] The purge, which Twitter characterized as a "routine" action against accounts "engaging in a mix of spamming, inauthentic behavior, and ban evasion," removed accounts

111 *See, e.g.*, *Annual Reports*, FACEBOOK INVESTOR RELATIONS, https://investor.fb.com/financials/default.aspx (follow hyperlinks to corresponding Form 10-K for each fiscal year) (last visited Mar. 26, 2020).

112 *See, e.g.*, TWITTER, RETROSPECTIVE REVIEW: TWITTER, INC. AND THE 2018 MIDTERM ELECTIONS IN THE UNITED STATES 3 (Feb. 4, 2019), *available at* https://blog.twitter.com/content/dam/blog-twitter/official/en_us/company/2019/2018-retrospective-review.pdf (using the term "coordinated inauthentic behavior").

113 Vijaya Gadde & Kayvon Beykpour, *Setting the Record Straight on Shadow Banning*, TWITTER BLOG (July 26, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html.

114 *Id.*

115 *See also* Vijaya Gadde, *Confidence in Follower Counts*, TWITTER BLOG (July 11, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Confidence-in-Follower-Counts.html ("If we detect sudden changes in account behavior, we may lock the account and contact the owner to confirm they still have control of it. These sudden changes in behavior could include Tweeting a large volume of unsolicited replies or mentions, Tweeting misleading links, or if a large number of accounts block the account after mentioning them . . . .").

116 *See* Anthony Ha, *Twitter Takes Down a Large Number of Chinese-Language Accounts Ahead of the Tiananmen Square Anniversary*, TECHCRUNCH (June 1, 2019, 3:24 PM), https://techcrunch.com/2019/06/01/twitter-china-takedown/ (describing Twitter's efforts to suspend accounts before the thirtieth anniversary of the Tiananmen Square massacre).

belonging to prominent anti-Communist activists in the United States.[117] "Sometimes our routine actions catch false positives or we make errors," the company tweeted.[118]

Facebook has emphasized that its focus on behavior allows it to *avoid* regulating on the basis of content. For example, in October 2018, Facebook removed 559 Pages and 251 accounts for violating its rules against coordinated inauthentic behavior.[119] In a blog post disclosing the take-down, Facebook's head of cybersecurity policy said that:

> the 'news' stories or opinions these accounts and Pages share are often indistinguishable from legitimate political debate. That is why it's so important we look at these actors' *behavior*—such as whether they're using fake accounts or repeatedly posting spam—rather than their *content* when deciding which of these accounts, Pages or Groups to remove.[120]

---

117 *Id.* The episode prompted U.S. Senator Marco Rubio to accuse Twitter of acting as a Chinese government censor. Marco Rubio (@marcorubio), TWITTER (June 1, 2019, 5:58 AM), https://twitter.com/marcorubio/status/1134806381775806464.

118 Twitter Public Policy (@policy), TWITTER (June 1, 2019, 7:16 AM), https://twitter.com/Policy/status/1134825963089465. In October 2018, Twitter pulled down "a network of suspected Twitter bots" for authenticity violations, after an NBC News journalist showed Twitter that the accounts, from Saudi Arabia, were tweeting and retweeting a set of pro-government talking points. Ben Collins & Shoshana Wodinsky, *Twitter Pulls Down Bot Network that Pushed Pro-Saudi Talking Points about Disappeared Journalist*, NBC NEWS (Oct. 18, 2018, 6:39 PM), https://www.nbcnews.com/tech/tech-news/exclusive-twitter-pulls-down-bot-network-pushing-pro-saudi-talking-n921871. The reporter discovered the bot network "by analyzing a trove of Twitter data and finding accounts that were created on the same date and had similar numbers of followers, tweets and likes. From there, he compiled a list of hundreds of accounts that tweeted identical tweets at the same time." *Id.* There was good evidence that this was a bot network: many accounts had been created on the same day, had similar numbers of followers, and tweeted at the same time. Notably, though, the content of the tweets was treated as evidence of inauthenticity. Months later, when "about 200,000" Saudi Arabian Twitter users defected to a new social media network, Parler, to protest the takedowns, a Reuters analysis presented the defectors as real people—political nationalists and supporters of Crown Prince Mohammed bin Salman—who objected to unexplained takedowns by the company. Elizabeth Culliford & Katie Paul, *Unhappy with Twitter, Thousands of Saudis Crash Pro-Trump Social Network Parler*, REUTERS (June 13, 2019, 4:37 PM), https://www.reuters.com/article/us-twitter-saudi-politics/unhappy-at-twitter-thousands-of-saudis-crash-pro-trump-social-network-idUSKCN1TE32S. The Reuters journalists interviewed a defector, who objected to Twitter banning accounts of nationalists "without explanation." *Id.* The story suggests that some accounts shut down for authenticity violations *were* fake accounts, but that others might not have been. If some real speakers were shut down as part of an authenticity enforcement exercise, it's likely because they tweeted or retweeted similar content as the "bad actors" the company sought to silence.

119 Nathaniel Gleicher & Oscar Rodriguez, *Removing Additional Inauthentic Activity from Facebook*, FACEBOOK NEWSROOM (Oct. 11, 2018), https://newsroom.fb.com/news/2018/10/removing-inauthentic-activity/.

120 *Id.* (emphasis in the original). Just a few days before the November 2018 midterm elections, Facebook pulled more fake accounts, claiming to have received a tip from U.S. law enforcement. Nathaniel Gleicher, *Election Update*, FACEBOOK NEWSROOM (Nov. 5, 2018),

In December 2018, Facebook published a video in which its head of cybersecurity policy explained more about "coordinated inauthentic behavior."[121] The official stated: "[w]hen we take down one of these [coordinated inauthentic] networks, it's because of their deceptive behavior. It's not because of the content they're sharing. The posts themselves may not be false, and may not go against our community standards."[122]

That month, Facebook pulled down the accounts of five American technology experts for "coordinated inauthentic behavior" in connection with their activities around the Alabama Senate election in 2017.[123] The technology experts had not employed false identities; rather, they had created a Facebook Page, titled "Alabama Conservative Politics," when they really "leaned Democratic."[124] The Page transmitted links to news articles published by major news outlets, such as the *Washington Post* and Fox News, and encouraged conservative voters to cast their ballots for a write-in candidate rather than the Republican nominee.[125] Facebook apparently

---

https://newsroom.fb.com/news/2018/11/election-update/ ("On Sunday evening, US law enforcement contacted us about online activity that they recently discovered and which they believe may be linked to foreign entities.").

121 Nathaniel Gleicher, *Coordinated Inauthentic Behavior Explained*, FACEBOOK NEWSROOM (Dec. 6, 2018), https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/.

122 *Id*. (video remarks of Nathaniel Gleicher).

123 *See, e.g.*, Scott Shane, *Facebook Closes 5 Accounts That Adopted Russian Tactics*, N.Y. TIMES, Dec. 23, 2018, at A27 (describing Facebook's actions of removing accounts associated with election deception in Alabama's U.S. Senate election). Only one of the five individuals has been identified: Jonathon Morgan, the Chief Executive Officer of an Austin-based cybersecurity firm, New Knowledge. According to the *New York Times*, Morgan has acknowledged participating in a "secret Alabama operation on Facebook and Twitter" but described it as "a small experiment designed to understand such techniques." *Id*. Democrat Doug Jones won the election.

124 *Id.* (explaining that the five "created a Facebook page on which they posed as conservative Alabamians"). In a twist, the deceptive campaign was reportedly funded by Reid Hoffman, the co-founder of another social media company, LinkedIn. *Id.*

125 *Id.* (describing the acts of the five technology experts as "deceptive tactics"); *see also* Craig Timberg, Tony Romm & Aaron C. Davis, *Researcher Whose Firm Wrote Report on Russian Interference Used Questionable Online Tactics During Alabama Senate Race*, WASH. POST (Dec. 18, 2018, 10:21 PM), https://www.washingtonpost.com/technology/2018/12/19/researcher-affiliated-with-russian-interference-senate-report-used-questionable-online-tactics-during-alabama-senate-race/ (explaining how a leading social media researcher, Jonathon Morgan, engaged in misleading online tactics during the Alabama election); Scott Shane & Alan Blinder, *Secret Experiment in Alabama Senate Race Imitated Russian Tactics*, N.Y. TIMES (Dec. 19, 2018), https://www.nytimes.com/2018/12/19/us/alabama-senate-roy-jones-russia.html (mentioning Morgan's involvement in the Alabama election). The deceptive tactics involved "creating a misleading Facebook page to appeal to conservatives" and "purchasing retweets on Twitter to measure the potential 'lift' of political messages." Tony Romm & Craig Timberg, *Facebook Suspends Five Accounts, Including That of a Social Media Researcher, For Misleading Tactics in Alabama Election*, WASH. POST (Dec. 22, 2018, 9:22 AM),

treated this as a violation of its rules against coordinated inauthentic behavior on the ground that it involved deception—a sort of political valence deception.[126]

Before the 2018 midterm elections, Twitter put together a "cross-functional team" devoted to "site and service integrity."[127] The team created a "political conversations dashboard" whose purpose was to "surface information about sudden shifts in sentiment around a specific conversation, suggesting a potential coordinated campaign of activity."[128] Twitter was studying political content on its network to uncover evidence of inauthenticity. It had put speakers' associations and political expression under the microscope in order to ferret out "bad faith" discourse. One can imagine how this approach might be useful for flagging a false news story or identifying the originator of a false news story—except that Twitter does not remove content for being false.[129] Rather, it looks for patterns in associations (and possibly changes in political valence) to establish inauthenticity, which *is* a basis for removing content.

C. Identity Verification

Facebook had a small program of identity verification dating back years before the 2016 election; the company originally began verifying identities when user reports challenged an account's authenticity. Starting in 2012, Facebook began offering identity verification more broadly, to certain public figures and celebrities, on a voluntary basis.[130] The company selected which

---

https://www.washingtonpost.com/technology/2018/12/22/facebook-suspends-five-accounts-including-social-media-researcher-misleading-tactics-alabama-election/.

126 Social media companies have long claimed the ability to infer users' political affiliations from their online behavior. *See, e.g.*, Jeremy B. Merrill, *Liberal, Moderate or Conservative? See How Facebook Labels You*, N.Y. TIMES (Aug. 23, 2016), https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html (noting how Facebook categorizes users into political affiliations based on liked pages). The use of such predictions to establish "inauthentic behavior" could result in a company flagging speech that falls outside the company's assessment of a speaker's political valence.

127 TWITTER, RETROSPECTIVE REVIEW: TWITTER, INC. AND THE 2018 MIDTERM ELECTIONS IN THE UNITED STATES 8 (Feb. 4, 2019), *available at* https://blog.twitter.com/content/dam/blog-twitter/official/en_us/company/2019/2018-retrospective-review.pdf

128 *Id.*

129 On March 27, 2020, Twitter published some new guidelines that suggest it is removing false content related to COVID-19. *See An Update on Our Content Moderation Work*, TWITTER BLOG (Mar. 27, 2020), https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#moderation (requiring people to remove tweets that include "[d]enial of established scientific facts about transmission" of COVID-19).

130 *See* Josh Constine, *Facebook Launches Verified Accounts and Pseudonyms*, TECHCRUNCH (Feb. 15, 2012, 10:07 PM), https://techcrunch.com/2012/02/15/facebook-verified-accounts-alternate-names/

celebrities and public figures were eligible for verification and, as an inducement, promised to promote verified users more frequently in the "subscribe suggestions" of the general public, which would likely increase their subscriptions.[131] Thus, in an early incarnation, identity verification involved a value-based exchange: the user turned over his or her data for verification in return for Facebook's promotion of the user to others. Today, verification still "comes with perks."[132]

Facebook did not introduce the blue checkmark as a visual confirmation of verification until 2013.[133] After the 2016 election, some critics argued that Facebook had contributed to toxic political discourse by giving "coveted blue check marks" "to partisan accounts on the right and left," thereby "lending them an air of credibility."[134]

("[Verified accounts are] a way to ensure people don't subscribe to the public updates of imposters."). Facebook had experienced some high-profile instances of stolen identity in the years preceding this move. In 2010, for example, the Facebook identity of the secretary general of Interpol, Ronald K. Noble, was impersonated by two different profiles, at least one of which was seeking to obtain information on targets of an Interpol operation. Josh Halliday, *Facebook Brings in Extra Safeguards to Block Scams*, GUARDIAN, Sept. 21, 2010, at 7. Writing in 2014, Chris Cox, a Facebook executive, explained that when a user account was reported as potentially fake, the company would "ask the flagged accounts to verify they are using real names by submitting some form of ID—gym membership, library card, or piece of mail." Chris Cox, FACEBOOK (Oct. 1, 2014, 2:22 PM), https://www.facebook.com/chris.cox/posts/10101301777354543. Twitter, a Facebook competitor, offered identity verification as early as 2009, but endured a scandal in 2012 when it erroneously verified a fake account as belonging to Wendi Deng, wife of News Corp.'s chairman Rupert Murdoch. *See* Adam Clark Estes, *How Twitter Accidentally Verified the Wrong Wendi Deng*, ATLANTIC (Jan. 4, 2012), https://www.theatlantic.com/technology/archive/2012/01/how-twitter-accidentally-verified-wrong-wendi-deng/333529/ (noting that the error was due to misplaced punctuation).

131 Josh Constine, *Facebook Launches Verified Accounts and Pseudonyms*, TECHCRUNCH (Feb. 15, 2012, 10:07 PM), https://techcrunch.com/2012/02/15/facebook-verified-accounts-alternate-names/ ("Stefani Germanotta, aka Lady Gaga, could use Verified Accounts to verify that she is the famous Stefani Germanotta, to display her name as 'Stefani Germanotta (Lady Gaga)', or display it as simply 'Lady Gaga' with Stefani Germanotta appearing in the About page of her profile. Lady Gaga would then appear more frequently in Facebook's Subscribe suggestions.").

132 Taylor Lorenz, *The Problem With Verification*, ATLANTIC (June 25, 2019), https://www.theatlantic.com/technology/archive/2019/06/instagram-and-twitter-should-eliminate-verification/592351/ ("[When verified,] your comments are sometimes featured higher, it's harder to impersonate you, and you get more robust insights on your personal account.").

133 Brittany Darwell, *Facebook Launches Verified Pages and Profiles to Help Users Identify Authentic Accounts*, ADWEEK (May 29, 2013), https://www.adweek.com/digital/facebook-launches-verified-pages-and-profiles-to-help-users-identify-authentic-accounts/.

134 Emma Roller, *Your Facts or Mine?*, N.Y. TIMES (Oct. 25, 2016), https://www.nytimes.com/2016/10/25/opinion/campaign-stops/your-facts-or-mine.html; *see also* Ariana Tobin, Madeleine Varner & Julia Angwin, *Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up*, PROPUBLICA (Dec. 28, 2017, 5:53 PM), https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes (noting that Facebook pages run by organizations

In April 2018, Facebook introduced new, mandatory identity verification for political advertisements.[135] It presented the program as a solution to problems of political advertisement transparency, which had become a focal point for social media critics after the 2016 election. Facebook's roll out emphasized the company's good-faith corporate citizenship in adopting the new measures, which, it noted, went beyond the requirements of campaign-finance law.[136]

The new rules applied to "U.S. advertisers and advertisers targeting the U.S." who sought to use Facebook's paid tools to enhance the distribution of their speech relating "to any national legislative issue of public importance in any place where the ad is being run."[137] Facebook provided a lengthy list of topics it considered de facto "national issues of public importance" in the United States, which included abortion, crime, health, and "values."[138] Facebook has used the term "issue ad" to describe communications covered by the rules, picking up on a term commonly used by election lawyers.[139] The topics requiring Facebook's authenticity pre-clearance reflect value judgments by the company; for example, "poverty" was a "political" topic under Facebook's rules but "wealth" was not.[140]

---

identified by the Southern Poverty Law Center as hate groups "are decked out with verification checkmarks").

135 *See* Rob Goldman & Alex Himel, *Making Ads and Pages More Transparent*, FACEBOOK NEWSROOM (Apr. 6, 2018), https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/ (preventing advertisers from running political advertisements until they are authorized).

136 *See, e.g.*, Mark Zuckerberg, *Preparing for Elections*, FACEBOOK NOTES (September 12, 2018), https://www.facebook.com/notes/mark-zuckerberg/preparing-for-elections/10156300047606634/ ("Facebook now has a higher standard of ads transparency than has ever existed with TV or newspaper ads.").

137 *Ads About Social Issues, Elections or Politics*, FACEBOOK BUSINESS, https://www.facebook.com/business/help/1838453822893854 (last visited Apr. 1, 2020). The full list was: abortion, budget, civil rights, crime, economy, education, energy, environment, foreign policy, government reform, guns, health, immigration, infrastructure, military, poverty, social security, taxes, terrorism, and values. The identity verification process was announced on April 23, 2018. *The Authorization Process for US Advertisers to Run Political Ads on Facebook is Now Open*, FACEBOOK BUSINESS (Apr. 23, 2018), https://www.facebook.com/business/news/the-authorization-process-for-us-advertisers-to-run-political-ads-on-facebook-is-now-open.

138 *Ads About Social Issues, Elections or Politics*, FACEBOOK BUSINESS, https://www.facebook.com/business/help/1838453822893854 (last visited Apr. 1, 2020).

139 *See* Rob Goldman & Alex Himel, *Making Ads and Pages More Transparent*, FACEBOOK NEWSROOM (Apr. 6, 2018), https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/ (noting that authorization will be required of "anyone that wants to show 'issue ads'—like political topics that are being debated across the country").

140 *See id.*

Under the rules, in order to use Facebook's paid tools to speak on a "political" subject, the speaker must "be authorized" by Facebook—a process that requires the speaker to "go through the U.S. residency and ID verification flow."[141] The speaker must submit to Facebook an image of his or her U.S. passport, driver's license, or state photo identification; at one time, a speaker was also required to submit the last four digits of his or her Social Security number. Facebook then mails a code to the speaker's residential address in the United States, which the user must input to confirm his or her identity. In other words, in the verification process, Facebook collects at least the following information about a user: a photo of his or her face, his or her full legal name, height and weight, date of birth, and home address. For some, the process may reveal a prior home address, information about visual and other impairments on a driver's license, and citizenship status, among other things. Thus, identity verification provides valuable, individualized data to Facebook about users. In essence, identity verification functions as a system of prior restraint, requiring users to obtain what amounts to a Facebook "license" in order to engage in certain types of political expression.[142]

Facebook has announced expansions of identity verification at least twice. In May 2018, Facebook clarified that it would apply its identity verification rules to all publishers of paid content, including news publishers.[143] In other words, news outlets that use Facebook's paid tools to boost content must put forward a business manager to go through identity verification. Since this announcement, Facebook has silenced paid posts promoting news stories on public policy matters by publications like the *Wall Street Journal*.[144]

---

141 *Ads About Social Issues, Elections or Politics*, FACEBOOK BUSINESS, https://www.facebook.com/business/help/1838453822893854 (last visited May 25, 2020); *Confirm Your Identity*, FACEBOOK BUSINESS, https://www.facebook.com/business/help/2992964394067299?id=288762101909005 (last visited May 25, 2020).

142 *See* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1177–78 (2018) ("[N]ew school speech regulation—which encourages blocking and filtering—is analogous to prior restraint.").

143 *See* Josh Constine, *Facebook and Instagram Launch US Political Ad Labeling and Archive*, TECHCRUNCH (May 24, 2018, 2:01 PM), https://techcrunch.com/2018/05/24/facebook-political-ad-archive/ (requiring "paid for by" labels on political and issue ads).

144 *See* Jeremy B. Merrill & Ariana Tobin, *Facebook's Screening for Political Ads Nabs News Sites Instead of Politicians*, PROPUBLICA (June 15, 2018, 12:39 PM), https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political (noting that "hot-button issues" are likely to pop up in posts from news organizations in addition to political ads, which creates complications).

Separately, Facebook has announced that it plans to expand identity verification beyond paid speech on political subjects. It has said that in the future it will require "people who manage Pages with large numbers of followers" to have their identities verified.[145] The company has made it clear that unverified individuals who manage "large Pages" "will no longer be able to post."[146] Instagram, which is owned by Facebook, introduced verification for a subset of users in 2014 and expanded its reach in 2018.[147]

Since their implementation, Facebook's identity verification rules have been repeatedly criticized for censoring legitimate political expression, as well as non-political speech.[148] An investigation by the *Washington Post* turned up numerous instances in which Facebook had refused to transmit paid content that included references to identity groups, such as promotions of events related to LGBT issues, without identity verification, because Facebook considered such content "political."[149] In particular, the *Washington Post* recounted the experience of Thomas Garguilo, a New Yorker who sought unsuccessfully to pay to promote a Facebook post about a panel discussion with an LGBT radio station in Washington. A Facebook employee responded to Garguilo's complaint by explaining that his proposed content "mentions LGBT which would fall under the category of civil rights

145 Rob Goldman & Alex Himel, *Making Ads and Pages More Transparent*, FACEBOOK NEWSROOM (Apr. 6, 2018), https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/; *see also* Testimony of Sheryl Sandberg, *supra* note 48, at 2 (stating that Facebook will require "people that run Pages with large audiences in the U.S." to "go through an authorization process and confirm their location").

146 Rob Goldman & Alex Himel, *Making Ads and Pages More Transparent*, FACEBOOK NEWSROOM (Apr. 6, 2018), https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/ ("[T]his will make it much harder for people to administer a Page using a fake account, which is strictly against our policies.").

147 *See* Taylor Lorenz, *The Problem with Verification*, ATLANTIC (June 25, 2019), https://www.theatlantic.com/technology/archive/2019/06/instagram-and-twitter-should-eliminate-verification/592351/ (noting expansion through the introduction of a public verification request form).

148 *See, e.g.*, David Gale, *Facebook's Problems with Veterans*, WALL STREET J. (Aug. 7, 2018, 6:55 PM), https://www.wsj.com/articles/facebooks-problem-with-veterans-1533682511 (quoting founder of We Are Mighty, a "media brand" for American military veterans and their families, objecting that under Facebook's new rules, "if anything in our posts uses the word 'military,' we are classified as a 'political' advertiser" and must register).

149 *See* Eli Rosenberg, *Facebook Blocked Many Gay-Themed Ads as Part of Its New Advertising Policy, Angering LGBT Groups*, WASH. POST (Oct. 3, 2018, 4:44 PM), https://www.washingtonpost.com/technology/2018/10/03/facebook-blocked-many-gay-themed-ads-part-its-new-advertising-policy-angering-lgbt-groups/ (noting that Facebook blocked dozens of advertisements mentioning LGBT themes and words).

which is a political topic."[150] The *Washington Post* reporters found that, in enforcing its identity verification rules, Facebook rejected advertisements that included, among other things, "a celebration of Nigerian Independence Day in Houston," "a post with facts about Holocaust diarist Anne Frank," and "a list of senior-friendly housing options distributed by a nonprofit group in Texas."[151]

Since 2009, Twitter has also employed verification; like Facebook, it originally offered a "blue badge" only to celebrities and public figures.[152] In 2017, Twitter temporarily suspended verification in response to public

150 *See id.* ("Garguilo said that so many of his ads have gotten blocked by Facebook that he has stopped using the words "LGBT" or "gay" in his language on the service.").

151 *Id.* Numerous other critics and investigators found that Facebook's identity verification rules were shutting down speech. ProPublica found that Facebook had refused, without identity verification, to promote many news articles published by independent news outlets, such as The Hechinger Report, Voice of Monterey Bay, and BirminghamWatch. Jeremy B. Merrill & Ariana Tobin, *Facebook's Screening for Political Ads Nabs News Sites Instead of Politicians*, PROPUBLICA (June 15, 2018, 12:39 PM), https://www.propublica.org/article/facebook-new-screening-system-flags-the-wrong-ads-as-political. The Nuyorican Poets Café, a New York cultural nonprofit, was prevented from paying to promote a post encouraging people to vote in the midterm elections. Daniel Gallant, *Facebook Censors at Random*, WALL STREET J. (Dec. 9, 2018, 5:52 PM), https://www.wsj.com/articles/facebook-censors-at-random-1544395970. Facebook stopped the Boston Museum of Fine Arts from paying to promote a lecture about sculpture. *Id.* Facebook prevented Arts Japan 2020, a cultural organization, from paying to promote a post "celebrating an award given by the emperor of Japan to an American arts curator." *Id.* The company stopped a nonprofit from advertising a fundraiser for disabled veterans. *See* J. Nathan Matias, Austin Hounsel & Melissa Hopkins, *We Tested Facebook's Ad Screeners and Some Were Too Strict*, ATLANTIC (Nov. 2, 2018), https://www.theatlantic.com/technology/archive/2018/11/do-big-social-media-platforms-have-effective-ad-policies/574609/ ("Facebook ruled that the fundraiser was 'related to politics or issues of national importance' and needed special authorization."). And it prevented a Long Island nonprofit, the LGBT Network, from paying to promote the Long Island Pride Parade, a beach concert, a pride-themed night at a New York Mets baseball game, and an LGBT-youth prom. *See* Eli Rosenberg, *Facebook Blocked Many Gay-Themed Ads as Part of Its New Advertising Policy, Angering LGBT Groups*, WASH. POST (Oct. 3, 2018, 4:44 PM), https://www.washingtonpost.com/technology/2018/10/03/facebook-blocked-many-gay-themed-ads-part-its-new-advertising-policy-angering-lgbt-groups/ (reporting that "The *Washington Post* found dozens of advertisements mentioning LGBT themes and words that [Facebook] blocked for supposedly being political," and that Facebook told the *Post* that "the majority" of these were "in error"). Facebook prevented Marsha Bonner, a motivational LGBT speaker, from paying to promote an NAACP-sponsored conference about LGBTQ people of color. *Id.*

152 Laignee Baron, *Twitter Wants to Verify All Users as a Way to Prove Identity*, FORTUNE (Mar. 9, 2018), https://fortune.com/2018/03/09/twitter-verification-all-users/; *see* Kurt Wagner, *This is Why Everyone is Upset About Twitter's Blue Check Mark Verification Policy*, VOX (Nov. 9, 2017, 2:58 PM), https://www.vox.com/2017/11/9/16629796/twitter-halts-verification-white-supremacist-jason-kessler-policy-blue-check-mark (noting that, at the time, Twitter verified "all kinds of accounts it considers 'of public interest,' including celebrities, athletes and journalists").

outcry about its verification of accounts of prominent white supremacists.[153] Twitter resolved the controversy by rescinding the blue badge from the white supremacists' accounts, and ever since has used rescission occasionally to punish verified users for content violations.[154] In 2018, for example, it removed the verified blue check from Louis Farakhan's account after he published an anti-Semitic tweet.[155]

D. Micro-Targeting

Companies' back-office customization, described in Part I.A above, must be understood as a core part of their speech regulation. Essentially, many social media companies earn profits by charging advertisers to use their advertisement customization tools. The tools use data analytics to help advertisers target their speech to certain recipients, on the basis of those recipients' identifying characteristics and behavior. This practice is known as micro-targeting.[156]

ProPublica has published several important articles on Facebook's micro-targeting tools.[157] One of its key observations is that Facebook's

---

153 *See id.* (discussing criticism of Twitter for verifying the account of Jason Kessler, one of the organizers of the "Unite the Right" rally in Charlottesville, Virginia, in 2017); Twitter Support (@TwitterSupport), TWITTER (Nov. 9, 2017, 8:03 PM), https://twitter.com/twitter support/status/928654369771356162? ("Verification was meant to authenticate identity & voice but it is interpreted as an endorsement or an indicator of importance. We recognize that we have created this confusion and need to resolve it. We have paused all general verifications while we work and will report back soon.").

154 Kurt Wagner, *This is Why Everyone is Upset About Twitter's Blue Check Mark Verification Policy*, VOX (Nov. 9, 2017, 2:58 PM), https://www.vox.com/2017/11/9/16629796/twitter-halts-verification-white-supremacist-jason-kessler-policy-blue-check-mark (noting that Twitter rescinded blue check marks from Jason Kessler and Richard Spencer).

155 Megan Keller, *Twitter Says It Won't Suspend Louis Farrakhan Over Tweet Comparing Jews to Termites*, HILL (Oct. 17, 2018, 6:24 PM), https://thehill.com/policy/technology/411950-twitter-says-it-wont-suspend-louis-farrakhan-over-tweet-comparing-jews-to (reporting that Twitter declined to shut down Farrakhan's account for anti-Semitic content, but had previously removed his verified status for a similar offense).

156 *See* Sonia Katyal, *Private Accountability in the Age of AI*, 66 UCLA. L. REV. 54, 91 (2019) ("Since websites often rely on predictive algorithms to analyze people's online activities . . . they can create profiles based on user behavior, and predict a host of identity characteristics that marketers can then use to decide the listings that a user sees online.").

157 *See, e.g.*, Julia Angwin, Madeleine Varner & Ariana Tobin, *Facebook Enabled Advertisers to Reach 'Jew Haters,* PROPUBLICA (Sept. 14, 2017, 4:00 PM), https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters (noting how Facebook's algorithm allowed anti-Semitic advertisement categories); Julia Angwin & Terry Parris Jr., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016, 1:00 PM), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race (noting that Facebook allows advertisements that exclude "groups it calls Ethnic Affinities.").

advertisement-targeting relies partly on identifying information that users voluntarily give Facebook, such as their age, gender, and location, and partly on other information, which is gleaned in ways that are likely not well-understood by users. These include the user's online behavior, such as his or her actions to associate (or disassociate) with other users, and the content of the user's expression on Facebook. So, for example, when a user "likes" something or posts content on Facebook, those acts are mined by Facebook to produce data that can be used as the basis for micro-targeting. A ProPublica study revealed that Facebook allowed "detailed targeting" of an advertisement audience based on categories such as "Demographics > Education > Field of study," which, until ProPublica published its findings, included subfields like "Jew hater."[158] The detail and granularity of Facebook's advertisement-targeting helps set it apart from competitors. Not only is Facebook's user base huge—in June 2019, the company had 244 million monthly active users in the United States and Canada alone[159]—but its micro-targeting tools are finely-tailored to individuals' identities and interests.

Following ProPublica's initial reporting, which raised concerns about racism and discrimination in advertisement-targeting, Facebook announced that it would disable "exclusion options" based on "ethnic affinities" in its advertisement tools for housing, credit, and employment advertisements.[160] In July 2018, Facebook went further in a settlement agreement with Washington State, which had begun investigating whether Facebook's advertisement-targeting practices violated state laws prohibiting unfair and

158 *Id.* In 2018, fair-housing groups sued Facebook for violating federal law by allowing housing advertisers to engage in microtargeting that purposefully excluded families with children and "users with interests based on disability and national origin." Complaint at 2, National Fair Housing Alliance v. Facebook, Inc. (S.D.N.Y. Mar. 27, 2018) (No. 18 Civ. 2689).

159 FACEBOOK, FORM 10-Q FOR THE QUARTERLY PERIOD ENDED JUNE 30, 2019, at 28 (2019), *available at* https://s21.q4cdn.com/399680738/files/doc_financials/2019/FACEBOOK_INC_10Q_20190725.pdf.

160 *See* Press Release, Wash. State Office of the Att'y Gen., AG Ferguson Investigation Leads to Facebook Making Nationwide Changes to Prohibit Discriminatory Advertisements on Its Platform (July 24, 2018), *available at* https://www.atg.wa.gov/news/news-releases/ag-ferguson-investigation-leads-facebook-making-nationwide-changes-prohibit (noting how the report led to pressure from civil rights advocates); Erin Egan, *Improving Enforcement and Promoting Diversity: Updates to Ethnic Affinity Marketing*, FACEBOOK NEWSROOM (Nov. 11, 2016), https://newsroom.fb.com/news/2016/11/updates-to-ethnic-affinity-marketing/ (noting Facebook's voluntary commitment to "[b]uild tools to detect and automatically disable the use of ethnic affinity marketing for certain types of ads," to update its Advertising Policies, and to require advertisers to affirm that they will not engage in discriminatory advertising).

discriminatory practices.[161] In the settlement, Facebook agreed to not provide an option for advertisers to limit the audience for certain types of advertisements by excluding some protected categories, such as race or veteran status.[162] Although Facebook agreed to extend this commitment across the United States, it remains free to earn fees for ad-targeting on the basis of other identifying characteristics, such as age, political affiliation, education, income, location, job, and health. Advertisements for other types of products and services can continue to target or exclude based on protected characteristics.[163]

---

161 Washington's investigation, begun in November 2016, focused on: section 19.86.020 of the *Washington Revised Code*, prohibiting "[u]nfair methods of competition and unfair or deceptive acts or practices in the conduct of any trade or commerce"; and section 49.60.030.1 of the *Washington Revised Code*, preserving rights to "be free from discrimination because of race, creed, color, national origin, sex, honorably discharged veteran or military status, sexual orientation, or the presence of any sensory, mental, or physical disability" in connection with employment, public accommodations, real estate transactions, credit transactions, and insurance transactions. Assurance of Discontinuance at 1, *In re* Facebook, Inc., No. 18-2-18287-5 SEA (Wash. King Cty. Super. Ct. July 24, 2018), *available at* https://agportal-s3bucket.s3.amazonaws.com /uploadedfiles/Another/News/Press_Releases/2018_07_23%20AOD.pdf

162 The types of advertisements were: employment advertisements, credit advertisements, insurance advertisements, and public accommodations advertisements. *See id.* at 4–5 ("[B]y way of example, Facebook would not allow the option within the Exclusion Targeting Tool to limit audiences based on a targeting category for 'Chinese people' or "Wheelchair users' because these categories, on their face, act as direct descriptors of Protected Characteristics. However, Facebook would not remove targeting categories such as 'Chinese literature' or 'Disability rights' as those categories identify interests and do not describe Protected Characteristics.").

163 Political advertisement targeting, in particular, has attracted commentators' attention. No public laws restrict micro-targeting of political advertisements, and Facebook's settlement with Washington State did not address the subject. In the lead-up to the 2016 election, the IRA and affiliated organizations spent about $100,000 on 3519 advertisements on Facebook and Instagram. In sworn testimony to the Senate Select Committee on Intelligence in September 2018, Sheryl Sandberg stated that the IRA "used coordinated networks of fake Pages and accounts to interfere in the election." Testimony of Sheryl Sandberg, *supra* note 48, at 2. As of May 2018, the advertisements themselves are available online. *Exposing Russia's Efforts to Sow Discord Online: The Internet Research Agency and Advertisements*, U.S. HOUSE OF REPRESENTATIVES PERMANENT SELECT COMMITTEE ON INTELLIGENCE, https://intelligence.house.gov/social-media-content/default.aspx (last visited Apr. 1, 2020); *see also* Deepa Seetharaman, Georgia Wells & Byron Tau, *Release of Thousands of Russia-Linked Facebook Ads Shows How Propaganda Sharpened*, WALL STREET J. (May 10, 2018, 12:12 PM), https://www.wsj.com/articles/full-stock-of-russia-linked-facebook-ads-shows-how-propaganda-sharpened-1525960804 (noting that the IRA accounts focused on racial and social issues early on, and as the election got closer the pages increasingly took on politics). Some of the IRA's advertisements targeted users in cities affected by racial unrest, such as Ferguson, Missouri; others targeted African-Americans, and some targeted users in swing voting states. Details of micro-targeting by the IRA is discussed in a series of *Wired Magazine* articles. *See* Issie Lapowsky, *House Democrats Release 3,500 Russia-Linked Facebook Ads*, WIRED (May 10, 2018, 10:00 AM), https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads/ (describing an advertisement targeted by the IRA at "users age 18 to 45 who were interested in BlackNews.com, the color black, or HuffPost Black Voices but were not Hispanic or Asian

E. Product Features

Social media companies find creative ways to leverage authenticity, verification, and users' data in new product features that shape political discourse. For example, Twitter announced in 2019 that it would suspend content-based restrictions for *some* public officials and candidates for public office, allowing only these individuals to publish tweets that violate content rules.[164] For a tweet to be eligible for this exemption, it must come from a *verified* official or candidate, and have more than 100,000 followers.[165]

In June 2017, Facebook introduced a "constituent badge" feature that permitted users to pin to their profiles an icon identifying them as constituents of particular elected officials.[166] The feature utilized the address the user attached to his or her account; Facebook only offered users a constituent badge for officials serving the districts in which they lived.[167] Facebook then offered elected officials the opportunity to communicate with tailored audiences of only verified constituents—for example, hosting "virtual town halls" on Facebook Live, attended only by verified constituents.[168] Journalists immediately pointed out that this system

American"); Issie Lapowsky, *How Russian Facebook Ads Divided and Targeted U.S. Voters Before the 2016 Election*, WIRED (April 16, 2018, 9:00 AM), https://www.wired.com/story/russian-facebook-ads-targeted-us-voters-before-2016-election/ (explaining that dark money advertisements and Russian-linked advertisements targeted voters in Pennsylvania, Virginia, and Wisconsin, and users in Wisconsin were "targeted with gun ads about 72 percent more often than the national average"). Trump campaign officials have claimed that the campaign used Facebook's paid micro-targeting tools to direct posts to "idealistic white liberals, young women, and African Americans," to discourage them from voting. *See* Joshua Green & Sasha Issenberg, *Inside the Trump Bunker, With Days to Go*, BLOOMBERG (Oct. 27, 2016, 6:00 AM) https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go ("We have three major voter suppression operations under way."). The article described how the Trump campaign used Facebook "dark posts" to send to "certain African American voters" an animated message with an anti-Clinton theme. *Id.*

164 *See Defining Public Interest on Twitter*, TWITTER BLOG (June 27, 2019), https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html (explaining that it is in the public's interest to have access to certain tweets).

165 *Id.* Twitter must also determine that the tweet is of public interest. *See id.* ("That said, there are cases, such as direct threats of violence or calls to commit violence against an individual, that are unlikely to be considered in the public interest.").

166 Griffin Connolly, *Facebook Features Connect Lawmakers With Constituents*, ROLL CALL (June 8, 2017, 5:01 AM), https://www.rollcall.com/politics/facebook-features-connect-lawmakers-constituents; *see also Tools for Government*, FACEBOOK FOR GOVERNMENT, POLITICS & ADVOCACY, https://politics.fb.com/tools-for-government/ (last visited Apr. 1, 2020) ("A constituent badge . . . appears next to [a person's] name when they comment on their elected official's Facebook post.").

167 Connolly, *supra* note 166.

168 *Id.*

significantly shaped news coverage of such events by, for example, preventing reporters from attending virtual town halls outside the districts in which they lived.[169] In its quest to "add value" to politicians' use of its network, Facebook had leveraged an aspect of its users' identities—the location of their homes—to govern the reach of their political speech.[170]

* * *

As outlined above, authenticity rules provide significant *business value* to social media companies. But social media companies do not typically justify authenticity regulation by pointing out its benefits to their business model. Instead, companies assert that authenticity is an important value that makes online discourse better, and that users should *be authentic* to further the important goal of improving free expression. The next Part turns to these claims and explores whether, in fact, authenticity is an important value that justifies companies' demands for accurate details about users' "true" selves as part of a broad, prosocial speech project.

II. THE VALUE OF AUTHENTICITY

Is authenticity an important expressive value? Or, does authenticity operate as a value that *limits* free speech, akin to the right of privacy? Or is authenticity, as the term is used by social media companies, something else altogether?

Recall that in Facebook's early days, Mark Zuckerberg justified the company's real-name policy by arguing that authenticity is a moral virtue.[171] Today, Facebook pairs "authenticity" with "integrity" in its Community Standards, reinforcing this moral ideal. In 2014, a Facebook executive offered two more justifications for Facebook's authenticity rules:

> First, it's part of what made Facebook special in the first place, by differentiating the service from the rest of the internet where pseudonymity, anonymity, or often random names were the social norm. Second, it's the

169 *Id.*

170 As part of its Town Hall project, Facebook also introduced "district targeting," which allows elected officials to create posts and polls that were visible only to confirmed constituents, and "constituent insights," which provided elected officials with tools to view and comment on news stories that were popular among their constituents. Faine Greenwood, *A Civics Lesson for Facebook*, SLATE (Aug. 8, 2017, 7:15 AM), https://slate.com/technology/2017/08/facebook-now-offers-constituent-services-what-could-go-wrong.html. All of these product features were designed to employ network technology and data analytics to create value for speakers or listeners by curating speech according to the identifying characteristics and behaviors of users.

171 *See* DAVID KIRKPATRICK, THE FACEBOOK EFFECT 199–200 (2010) ("Having two identities for yourself is an example of a lack of integrity.").

> primary mechanism we have to protect millions of people every day, all around the world, from real harm. The stories of mass impersonation, trolling, domestic abuse, and higher rates of bullying and intolerance are oftentimes the result of people hiding behind fake names, and it's both terrifying and sad.[172]

Four years later, a different Facebook executive explained that:

> Facebook was built for conversation and human connection. It's why we ask that people using our service be themselves—whether it's an individual, a business or a nonprofit. [W]e've worked hard to establish authenticity as a social norm because it's at the heart of most meaningful connections on Facebook.[173]

In 2019, Facebook amended its Community Standards to include an entry for "Authenticity": "We want to make sure the content people are seeing on Facebook is authentic. We believe that authenticity creates a better environment for sharing, and that's why we don't want people using Facebook to misrepresent who they are or what they're doing."[174]

Obviously, these justifications are all different. However, they advance a few key ideas: First, it is bad to present yourself as anything other than what you are. Second, when people have to reveal their "true selves," they are less likely to produce abusive or hateful speech. Third, authenticity is an essential component of "meaningful" expression, especially because it encourages users to express themselves (it "creates a better environment for sharing"), but also because authenticity produces "authentic content." A common theme in these ideas is that a speaker's authenticity is important mainly because it generates benefits for others—for the communicative community. In this view, speaker authenticity is an important value because it enhances discourse, and therefore is worth enforcing not only through informal social conventions but also sometimes through (private) regulation.

Yet this premise conflicts with modern ideas about the meaning of "authenticity." Disciplines ranging from philosophy[175] to social

---

172 Chris Cox, FACEBOOK (Oct. 1, 2014, 2:22 PM), https://www.facebook.com/chris.cox/posts/10101301777354543. It is not clear what Cox meant when he said that "domestic abuse" is caused by "people hiding behind fake names."

173 Alex Stamos, *Authenticity Matters: The IRA Has No Place on Facebook*, FACEBOOK NEWSROOM (Apr. 3, 2018), https://newsroom.fb.com/news/2018/04/authenticity-matters/.

174 *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/ (last visited Apr. 1, 2019); Monika Bickert, *Updating the Values That Inform Our Community Standards,* FACEBOOK NEWSROOM (Sept. 12, 2019), https://newsroom.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/.

175 The philosophical literature on authenticity reaches back at least as far as Aristotle's *Nicomachean Ethics*. *See generally* Somogy Varga & Charles Guignon, *Authenticity*, *in* STANFORD ENCYC. OF PHILOSOPHY 4 (Edward Zalta, Uri Nodelman & Colin Allen eds., 2017), *available at*

psychology[176] to pop psychology[177] have produced literatures on authenticity, mainly recognizing it as producing benefits for the self.[178] These disciplines tend to characterize authenticity as one or more of the following: (1) a process of *introspection and self-definition* (as in: "be true to yourself")[179]; (2) the act of *following one's heart* or one's intuition[180]; (3) *consistency* between one's values and

---

https://plato.stanford.edu/archives/fall2017/entries/authenticity/ (describing how an "older concept of sincerity, referring to being truthful in order to be honest in one's dealings with others," was eventually "replaced by a relatively new concept of authenticity, understood as being true to oneself for one's own benefit"); CHARLES TAYLOR, THE ETHICS OF AUTHENTICITY (1991).

176 *See, e.g.*, Susan Harter, *Authenticity*, *in* HANDBOOK OF POSITIVE PSYCHOLOGY (C.R. Snyder & Shane J. Lopez eds., 2002) (finding that adolescents' authenticity is correlated with psychological benefits); Alison P. Lenton et al., *How Does "Being Real" Feel? The Experience of State Authenticity*, 81 J. PERSONALITY 276, 285 (2013) (positing that authenticity may be precursor to positive affect); Leonard Reinecke & Sabine Trepte, *Authenticity and Well-Being on Social Network Sites: A Two-Wave Longitudinal Study on the Effects of Online Authenticity and the Positivity Bias in SNS Communication*, 30 COMPUTERS HUM. BEHAV. 95, 96 (2014) (suggesting that authenticity is a strong positive predictor of psychological health and well-being); Brenda K. Wiederhold, *Being Authentic on Facebook Has Same Health Benefits as In-Person Authentic Behavior*, 20 CYBERPSYCHOLOGY BEHAV. & SOC. NETWORKING 345, 345 (2017) (noting that "contentment, calmness, and social ease were common feelings when one is authentic").

177 *See, e.g.*, Christopher D. Connors, *The 5 Qualities of an Authentic Person*, MEDIUM (Feb. 28, 2017), https://medium.com/personal-growth/the-5-key-ingredients-of-an-authentic-person-259914abf6d5 (listing five ways to be authentic); Diane Mottl, *Ways of Living an Authentic Life*, PSYCH CENTRAL (Oct. 8, 2018), https://psychcentral.com/lib/ways-of-living-an-authentic-life/ (advising readers to critically determine what they truly believe in order to become authentic).

178 The psychology literature, in particular, has drawn connections between feelings of authenticity and feelings of individual well-being, and between authentic self-presentation on social media and feelings of well-being. *See, e.g.*, Lenton et al., *supra* note 176, at 285 ("[A] feeling of contentment and comfort with oneself and with others, when combined with a sense of one's own individuality (autonomy) and competence, are indicative of authenticity."); Reinecke & Trepte, *supra* note 176, at 96 (summarizing studies that had "found strong correlations between authenticity and various indicators of well-being such as self-esteem, positive affect, and satisfaction with life"); *id.* at 100 (proposing that "authentic self-presentation" on social media produced "positive effects on psychological well-being," but users with "lower levels of well-being" benefited less and struggled to present their "authentic negative feelings and experiences"); Wiederhold, *supra* note 176, at 345 ("[T]he field of positive psychology . . . confirms that being authentic correlates with higher levels of life satisfaction and well-being.").

179 From the psychology literature, see, e.g., Reinecke & Trepte, *supra* note 176, at 96 ("'[F]eeling real' and free of psychological tension between one's social roles and behavior and one's true self is the essence of the psychological concept of authenticity."). From the pop-psychology literature, see, e.g., Mottl, *supra* note 177 (arguing that authenticity "is being ourselves, not an imitation of what we think we should be or have been told we should be" and that "[b]eing authentic is more than being real; it is finding what is real").

180 From the psychology literature, *see, e.g.*, Lenton et al., *supra* note 176, at 277 ("Theorists from the humanistic tradition suggest that emotions are central to authenticity because a feeling of authenticity signals to the individual that the self is integrated and organized.") (citation omitted). From the pop-psychology literature, see, e.g., Connors, *supra* note 177 ("Listen to your intuition. Do what your heart tells you to do. That's what an authentic person does.").

one's acts, or "self-concordance"[181]; and (4) being *present and engaged* in the moment or unscripted.[182] In this form of authenticity, one's outward identity might not match one's internal identity. And, authenticity cannot be conferred by an outside party.[183] It is worth striving for because it is self-actualizing and *personal*.

The social media industry's "authenticity" is the opposite of this. According to the industry, authenticity means that you have revealed your one "true" identity by exposing only accurate personal details about yourself, for the purpose of benefiting the community. It is *not* about *your* self-actualization or self-expression.[184] And whether or not you *are* authentic will be judged by a third-party corporation. A graphic embedded in Facebook's

---

181 From the psychology literature, see, e.g., Lenton et al., *supra* note 176, at 277 (finding that trait-based and state-based authenticity have been "operationalized similarly" as "value- or trait-behavior consistency," sometimes labeled "self-concordance"). From the pop-psychology literature, see, e.g., Connors, *supra* note 177 ("There's never any doubt or questioning the integrity of an authentic individual. Their behavior, in terms of ethics and morals, is as predictable as snow during wintertime in Minnesota. You know what you're going to get."); Mottl, *supra* note 177 ("Being authentic . . . is when our actions and words are congruent with our beliefs and values.").

182 *See, e.g.*, Sand Farnia, *Snapchat and the Authenticity Revolution*, MEDIUM (Feb. 12, 2016), https://medium.com/start-up-vision/snapchat-and-the-authenticity-revolution-4cd3ecb8ef26 (describing how Snapchat felt "liberating" because "it mimicked real life, in that the moments came and went" and that as a result "content became more authentic. . . . Imperfection no longer mattered, because it was temporary, like memories"). Psychology research has found that the state of *feeling* authentic is most closely associated with contentment, calmness, enthusiasm, and love, while the feeling of inauthenticity ("feeling untrue") was associated with anxiety and public self-consciousness. *See* Lenton et al., *supra* note 176, at 286 (describing study finding that anxiety was the "signature emotion of least-me experiences" and "public self-consciousness was uniformly high").

183 *See* Lenton et al., *supra* note 176, at 286 (highlighting that "rejection of external influence" is typically considered a "criteria that must be fulfilled for an individual or a behavior to be deemed 'authentic'"). In the psychology literature, inauthenticity or false-self behavior has been described as "saying what you think *others* want to hear, not what you really think." Harter, *supra* note 176, at 382; *see also id.* at 383 (observing that "[t]ypically the false self is experienced as *socially implanted* against one's will").

184 Presenting your "true self" online should allow social media to better customize content for you, which the industry would argue is for your own benefit. On the other hand, if (as Zuboff contends) the end point of surveillance capitalism is not customization but prediction (or even manipulation), then it is not for your benefit.

Community Standards webpage visually depicts "Authenticity" as an expressionless woman shining a light on her own face while a figure watches:

Figure 1: "Authenticity" as Depicted by Facebook[185]

In 2019, for the first time, Facebook presented authenticity as a value that places *limits* on free expression, rather than one that benefits speech. In the preamble of the charter for its new Oversight Board, Facebook asserted that, at times, "speech can be at odds with authenticity, safety, privacy, and dignity."[186] This implied that authenticity is a separate, independent ethic or right, analogous to safety, privacy, and dignity.[187] Unlike authenticity, however, these other things are *individual* rights—basic human rights, in fact.[188] Authenticity is not generally considered a human right on par with safety, privacy, or dignity.[189] This may be because, in the conventional understanding, authenticity comes from within, and must be secured by each individual for him- or herself.[190] The statement in Facebook's Oversight

185 *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/ (last visited Apr. 1, 2020).

186 OVERSIGHT BOARD CHARTER, *supra* note 37, at 2.

187 The implication was made clearer in a letter by Mark Zuckerberg, which Facebook posted online with the Charter, to "explain[] the board's purpose and goals." Brent Harris, *Establishing Structure and Governance for an Independent Oversight Board*, FACEBOOK NEWSROOM (Sept. 17, 2019), https://newsroom.fb.com/news/2019/09/oversight-board-structure/. Zuckerberg wrote that when Facebook enforces its Community Standards, "we follow a set of values—authenticity, safety, privacy, and dignity—guided by international human rights standards." Mark Zuckerberg, *Facebook's Commitment to the Oversight Board*, FACEBOOK NEWSROOM (Sept. 17, 2019), https://fbnewsroomus.files.wordpress.com/2019/09/letter-from-mark-zuckerberg-on-oversight-board-charter.pdf. Although Zuckerberg's statement suggests that authenticity is a value connected to human rights, authenticity is *not* a widely recognized human-rights concept.

188 *See, e.g.*, G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948) (declaring safety (Article 3), privacy (Article 12), and dignity (Article 1) to be human rights).

189 Thus, for example, the Universal Declaration of Human Rights does not mention authenticity, though it does refer to obligations of the state to create conditions in which an individual's "full personality" may be "developed." *Id.* art. 29.

190 Note that insofar as the other values (safety, privacy, and dignity) operate to limit speech, they limit the speech of *others*. For example, the law might limit someone else's speech to preserve *my* safety. However, when authenticity operates to limit speech, it limits one's *own* speech. In order to preserve my authenticity, I (or a social media company) must limit my own (inauthentic) speech.

Board Charter offers yet a different view of authenticity—a reminder of how the industry's concept of authenticity is continually changing.[191]

This Part examines the social media industry's claim that speaker authenticity is an important "value"—morally, pragmatically, and expressively. It concludes that speaker authenticity *can* have unique value for online communication, where people never come face to face. But when "authenticity" is used in this sense, it generally refers to truthfulness about specific details of a person's identity that are relevant to the person's speech—not a broad obligation to present a single, "true" identity through myriad accurate data points. This Part finds little support for the claim that authenticity in the industry sense is morally virtuous. It finds mixed support for the claim that authenticity cuts down on abusive behaviors, with some research suggesting that abusive speech can *increase* when speakers are forced to disclose their identities. And it notes some reasons in favor of authenticity regulation that social media companies do not tend to bring up: not only the *business value* of authenticity, explored at length in Part I, but also that identity-based take-downs are necessary to prevent fraud and other crimes, such as foreign election interference, and are therefore essential to digital security. It discusses two other ways in which authenticity regulation benefits companies as business entities—by helping them avoid intrusive government regulation and protect their legal interests.

After considering these arguments, this Part considers some counterarguments. For example, authenticity regulation tends to legitimize speech from "authentic" speakers no matter its content, teaching that only speakers are bad, not ideas. Authenticity regulation treats anonymous or pseudonymous speech as if it has little or no value, which runs counter to long-held American free-speech commitments to pseudonymous political speech. Companies' efforts to treat inauthentic identity as a *behavior* may place significant "New Governor" scrutiny on individuals' *associations*. Micro-targeting shapes discourse without listeners' knowledge or consent. And there is an orwellian vagueness to the term "authenticity" when companies use it; unmoored from its socially constructed meaning (self-actualization), authenticity can mean anything. When content is treated as evidence of (in)authenticity, authenticity rules can operate as a form of quasi-content moderation.

---

191 The word "authenticity" appears only once in the Charter, in the quoted sentence. OVERSIGHT BOARD CHARTER, *supra* note 37, at 2.

A rigorous regulatory focus on authenticity encourages identity theft, creating ripple effects across the economy. In general, identity theft crimes have been increasing in the United States.[192] The increase in identity theft incentivizes companies to partner with law enforcement, a process that has been underway in earnest for years. The ensuing arms race—waged between sophisticated identity thieves and a state-private coalition—likely accelerates line-blurring between state surveillance and private surveillance, as well as state and private power.[193] This concern is related to the commodification of authentic identity, as verified by Facebook and other social media companies. Authenticity regulation makes an authentic identity a precious possession, and its verification a valuable service. Identity verification barely existed as a marketable service a few years ago; in the future, social media companies may generate significant value from this new line of business.

In sum, this Article disagrees that authentic speaker identity is a core value of free expression, at least as it is conceptualized by the social media industry. And it finds more risks than benefits to authenticity regulation, especially considering that content-based moderation can do much of the work necessary to keep networks secure from crime, misinformation, and so on. All this suggests that we should be skeptical of authenticity—as a moral precept, a speech value, and a stand-in for "authentic content"—as we move forward with online content moderation and speech regulation.

A. Bad Actors, Revisited

1. Is it Immoral to Disguise Your Identity?

In an interview, Mark Zuckerberg once emphatically repeated "[y]ou have one identity," three times in a single minute, impressing the interviewer with his moral zeal for the company's authenticity rules.[194] Of course, Zuckerberg spoke as a young, wealthy, white, healthy, cisgender,

192 *See* ERIKA HARRELL, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, NCJ 251147, VICTIMS OF IDENTITY THEFT, 2016, at 2 (2019), *available at* https://www.bjs.gov/content/pub/pdf/vit16.pdf (noting an increase in the prevalence of identity theft from 7% in 2014 to 10% in 2016).

193 Mark Zuckerberg has repeatedly referred to cybersecurity at Facebook as an "arms race." *See, e.g.*, *Facebook: Transparency and Use of Consumer Data: Hearing Before the H. Comm. on Energy & Commerce*, 105th Cong. 209 (2018) (statement of Mark Zuckerberg, Chairman and Chief Exec. Officer, Facebook), *available at* https://docs.house.gov/meetings/IF/IF00/20180411/108090/HHRG-115-IF00-Transcript-20180411.pdf ("Every problem around security is sort of an arms race, right?").

194 DAVID KIRKPATRICK, THE FACEBOOK EFFECT 199–200 (2010).

heterosexual, man from an upper-middle-class East Coast family, who had attended prestigious schools and was celebrated for his entrepreneurial success. Unlike many Facebook users, he possessed few identity characteristics that make people targets for abuse or discrimination.[195]

Zuckerberg's presentation of authenticity as a moral virtue was insensitive to the reasons that other people might want to hide aspects of their identity when they communicate online.[196] Those reasons abound. For example, research has shown that women and people of color—and particularly women of color—endure more abuse in online communication than men and white people.[197] For vulnerable speakers, rules that require them to be honest about their identities are a double-edged sword: in some speech settings, the disclosure might affirm a connection to a community, with positive effects on that person's self-expression, and on the community's conversation.[198] In other situations, however, the disclosure can affect listeners' evaluation of the person's speech (prompting listeners to treat it as less important or intelligent, for example), or mark the speaker for abuse. It seems reasonable for a speaker to try to avoid discrimination or abuse—or to try to have her expressive contributions taken equally as seriously as those

---

195 Zuckerberg is Jewish, and has been the target of anti-Semitic abuse, however. *See, e.g.*, *German Paper Ripped For 'Anti-Semitic' Caricature of Facebook's Zuckerberg*, FOX NEWS (Dec. 11, 2015), https://www.foxnews.com/world/german-paper-ripped-for-anti-semitic-caricature-of-facebooks-zuckerberg.

196 Indeed, something else may help explain Zuckerberg's passion for authenticity. Some psychological research has found that "simply being primed with power makes people feel more authentic." Lenton et al., *supra* note 176, at 286 (citing Michael W. Kraus, Serena Chen & Dacher Keltner, *The Power to Be Me: Power Elevates Self-Concept Consistency and Authenticity*, 47 J. EXPERIMENTAL SOC. PSYCHOL. 974 (2011)). Since Zuckerberg enjoyed significant social power—defined as "elevated control over others' outcomes and increased freedom to make decisions according to [his] own goals and motivations"—as a millionaire CEO, this research suggests that he may have felt greater-than-average "self-concept consistency" and authenticity. Michael W. Kraus, Serena Chen & Dacher Keltner, *The Power to Be Me: Power Elevates Self-Concept Consistency and Authenticity*, 47 J. EXPERIMENTAL SOC. PSYCHOL. 974, 974 (2011).

197 *See, e.g.*, Emma A. Jane, *Online Misogyny and Feminist Digilantism*, 30 CONTINUUM: J. MEDIA & CULTURAL STUD. 284, 284 (2016) ("[M]edia accounts and self-reports of sexualized electronic vitriol present a strong *prima facie* case that gendered cyber-hate has increased markedly since at least 2011."); *Toxic Twitter—A Toxic Place for Women*, AMNESTY INT'L, https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/ (pointing to Twitter as an example of a social platform where women of color endure abuse). *See generally* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014) (exploring the effect of cyber harassment, noting that it particularly affects women of color); Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373 (2009) (arguing that the gender discrimination law may help combat cyber gender harassment).

198 *See supra* Part II.A.2 (describing how online authenticity enables speakers to create targeted communities).

of other speakers—by disguising a component of her identity. And it seems *unfair*—and arguably even *immoral*—for social media networks to forbid vulnerable users from engaging in self-protection from these foreseeable harms.

Even putting aside discrimination and abuse, a person may be morally justified in obscuring aspects of her identity, or even deceiving others about her identity, if doing so preserves her privacy or dignity or allows her to better express her true self. If it is morally virtuous to "be authentic" for the benefit of others, this refers to being *truthful* toward others in your speech, which is the moral virtue of truthfulness. But social media companies do not want to present truthfulness as an important value for online speech, because they do not want to be "arbiters of truth."[199]

Social media companies that rely on authenticity enforcement offer little to users who, for various reasons, seek out online communication as a relief from the daily grind of in-person bias. And surely deceptive action is warranted to thwart a system which will attribute to you characteristics and qualities, such as political valence[200] or health status[201], against your wishes, to turn a profit for someone else. For all of these reasons, Zuckerberg's moral claims about authenticity fall short.

2. Authenticity and Trust

Social media companies assert that authenticity improves speech by encouraging meaningful connections through trust. This argument is more persuasive. There really *are* ways in which authenticity (in the industry sense) can enhance trust in communication. One is when part of a speaker's identity sheds light on the truthfulness of their speech. For example, if a person has never served in the military, the person can hardly claim to have

199 Mark Zuckerberg, FACEBOOK (Nov. 18, 2016, 12:15 AM), https://www.facebook.com/zuck/posts/10103269806149061?mod=article_inline ("We do not want to be arbiters of truth . . . .").

200 *See* Jeremy B. Merrill, *Liberal, Moderate, or Conservative? See How Facebook Labels You*, N.Y. TIMES (Aug. 23, 2016), https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html ("And now, it is easy to find out how Facebook has characterized you—as very liberal or very Conservative, or somewhere in between."); Michael D. Conover et al., *Predicting the Political Alignment of Twitter Users*, 3 IEEE INT'L CONF. ON SOC. COMPUTING 192, 198 (2011), *available at* http://www.bgoncalves.com/download/finish/4/53.html ("[P]olitically-active Twitters users generate text- and network-based information that can be used to effectively predict the political alignment of large numbers of individuals.").

201 *See, e.g.*, Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (describing how a company uses data about your health status to deploy targeted advertisements).

experienced something as a member of the military. If the person's lack of military service was known to their audience, it would affect how the audience evaluates the truthfulness of the person's speech.[202]

Online communication presents real challenges for speakers and listeners negotiating authenticity. In digital space, it can be difficult to confirm that you are communicating with someone who is who they say they are or to verify speakers' claims.[203] Catfishing, for example, is a widespread online practice in which one person tricks another about his or her identity, sometimes as part of a fraudulent scam.[204] Catfishing is mainly known for causing economic and even physical harm to victims, but it also causes expressive harm by causing people to mistrust online communication.

The question is whether companies' enforcement of authenticity really makes speech trustworthy by reducing false content and misrepresentations. Surely it must in some cases. When companies remove fake accounts, they prevent those accounts from spreading false information and committing fraud. However, social media websites are still filled with misrepresentations, exaggeration, and outright fraud. Authenticity regulation likely reduces this problem, but it has not proven particularly well-tailored to eliminate it.

Another reason that authenticity might matter to speech involves expressive communities. Some people derive significant expressive satisfaction from participating in online forums comprised of people who share something in common: working dads, for example, or struggling musicians. One of the unique benefits of social media networks is that they can bring together a group of geographically dispersed individuals who all share interests or characteristics. Facebook, for example, has numerous public Facebook groups on various subjects, and people enjoy these forums for expression.[205] In a community-based speech forum, participants may feel

---

202 In 2018, a Texas man gave interviews to the news media with a first-hand account of a school shooting. *Texas Man Said He Was a Survivor of the Santa Fe High School Shooting, He Was Lying*, NPR (July 3, 2019), https://www.npr.org/2019/07/03/738586985/texas-man-said-he-was-a-survivor-of-the-santa-fe-high-school-shooting-he-was-ly. The man said he was a substitute teacher at a high school where ten people were killed by a gunman. *Id.* But it turned out that the man had never worked for the school. *Id.* A reporter discovered the lie and debunked the man's story, revealing in the process that many news outlets had not verified important facts about the man's identity. *Id.*

203 It can also be challenging to negotiate authenticity in the real world. *See* United States v. Alvarez, 567 U.S. 709, 713 (2012) (analyzing a case involving a man who lied about being a famous hockey player, marrying a Mexican starlet, and receiving the Congressional Medal of Honor).

204 *See, e.g.*, Jack Nicas, *Facebook Connected Her to a Tattooed Soldier in Iraq. Or So She Thought.*, N.Y. TIMES (July 28, 2019), https://www.nytimes.com/2019/07/28/technology/facebook-military-scam.html.

205 Private Facebook groups introduce a moderator who gives permission for participants to join, and therefore, assumes the task of authenticating members.

freer to express themselves because they believe they are communicating only with people with whom they share something in common. And, they may credit the other participants' speech more on the basis of this commonality. Since one of the main benefits of the Internet is to foster this unique kind of connection across great distances—bringing the world together, in Facebook's motto[206]—we can say that authenticity offers real value for this kind of online communication.

But when we say that authenticity adds value to such communications, do we mean that speakers must present only accurate details about themselves online, with no deceptions (or obfuscations) about their "true" identity? Probably not. For example, if you belong to a Facebook group for working dads, it is probably important that you *really are* a working father. But is it important for others in the group to know your race, your zip code, or your educational background? No. Is it important for Facebook to know these things? Not for the specific purpose of ensuring that the working dads forum is limited to participants who really are working dads.

What all this suggests is that where authenticity has value for speech, it is really functioning as a stand-in for the *truthfulness* of specific claims. It is not necessary to be authentic in the industry sense—to reveal a wealth of accurate details about yourself for others to consume. Nor is it necessary to be authentic in the conventional sense—being true to yourself, consistent in your beliefs and actions, self-aware, and unscripted (i.e., not generally "phony"). It just matters if it is true that you are a working dad.

3. Authenticity and Anti-Social Behavior

Many participants in the technology industry believe that online anonymity facilitates "anti-social behavior," such as hate speech, harassment, and trolling.[207] In fact, little research has explored the

206 *See* Sarah Frier & Max Chafkin, *Zuckerberg's New Mission for Facebook: Bringing the World Closer*, BLOOMBERG (June 22, 2017), https://www.bloomberg.com/news/articles/2017-06-22/zuckerberg-s-new-mission-for-facebook-bringing-the-world-closer.

207 *See, e.g.*, Lee Rainie, Janna Anderson & Jonathan Albright, *The Future of Free Speech, Trolls, Anonymity and Fake News Online*, PEW RES. CTR. (Mar. 29, 2017), https://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/ (noting that "anonymity abets anti-social behavior" was a theme expressed by respondents in a 2016 survey of technology experts, scholars, corporate practitioners and government leaders); *The Twitter Paradox: How a Platform Designed for Free Speech Enables Internet Trolls*, NPR (Oct. 26, 2016), https://www.npr.org/2016/10/26/499442453/the-twitter-paradox-how-a-platform-designed-for-free-speech-enables-internet-tro ("Facebook has its own problems with abuse and harassment but not nearly to the same degree [as other social media networks] because there's no way for people

relationship between speaker anonymity and uncivil or aggressive online speech.[208] The evidence that exists is mixed, with some recent studies finding, contrary to the conventional wisdom, that the most aggressive social media speech comes from individuals operating under their real names.[209] For example, a 2014 study of a German social media platform found that commenters operating under their real names presented *more* online aggression than commenters whose identities were anonymous.[210] The researchers hypothesized that this might be because operating under their

---

to sort of hide behind an anonymous account name or an anonymous avatar"). *But see* Katja Rost, Lea Stahel & Bruno S. Frey, *Digital Social Norm Enforcement: Online Firestorms in Social Media*, 11 PLOS ONE, no. 6, June 7, 2016, at 1 (arguing that non-anonymous individuals are more aggressive in unleashing what the authors call "online firestorms" than their anonymous counterparts).

208 *See* Arthur D. Santana, *Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards*, 8 JOURNALISM PRAC. 18, 18 (2014) (noting "a striking dearth of empirical evidence in the academic literature of the effect that anonymity has on commenters' behavior" on online newspaper comment boards).

209 *See, e.g.*, Mikyeung Bae, *The Effects of Anonymity on Computer-Mediated Communication: The Case of Independent Versus Interdependent Self-Construal Influence*, 55 COMPUTERS HUM. BEHAV. 300, 305 (2016) (studying U.S. and South Korean undergraduates in online discussion forums and finding "no significant effects of anonymity on flaming or critical comments" among subjects and that "identifiable participants exhibited more critical comments than did anonymous participants,"—a finding that "is contrary to popular belief that anonymity enhances disinhibitive behavior"); Katja Rost, Lea Stahel & Bruno S. Frey, *Digital Social Norm Enforcement: Online Firestorms in Social Media*, 11 PLOS ONE, no. 6, June 7, 2016, at 1, 6, 18 (using a large dataset study of a German social media platform and finding that "more online aggression" was demonstrated by non-anonymous commenters, potentially because "[n]on-anonymity helps to gain recognition, increases one's persuasive power, and mobilizes followers"); *see also* Daegon Cho & K. Hazel Kwon, *The Impacts of Identity Verification and Disclosure of Social Cues on Flaming in Online User Comments*, 51 COMPUTERS HUM. BEHAV. 363 (2015) (demonstrating that policy-driven regulation that increases the likelihood users will be readily identified in online forums actually increases online animosity); Charlene Christie & Emily Dill, *Evaluating Peers in Cyberspace: The Impact of Anonymity*, 55 COMPUTERS HUM. BEHAV. 292, 292, 297 (2016) ("Only those participants with high self-esteem, low levels of social anxiousness, or an elevated sense of autonomy evaluated targets more negatively when anonymous rather than identifiable," while "the opposite pattern emerged among people with elevated social anxiousness"). *But see* Ian Rowe, *Civility 2.0: A Comparative Analysis of Incivility in Online Political Discussion*, 18 INFO., COMM. & SOC. 121, 121 (2015) (finding that political comments on the *Washington Post*'s website, where users enjoy a high level of anonymity, were more uncivil than responses to political news content on the *Washington Post*'s Facebook page, where there is less anonymity); Arthur D. Santana, *Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards*, 8 JOURNALISM PRAC. 18, 28 (2014) (finding that there is a dramatic improvement in civility when anonymity is removed). An excellent, though somewhat dated, review of the social science literature about the relationship between online behavior and anonymity is Kimberly M. Christopherson, *The Positive and Negative Implications of Anonymity in Internet Social Interactions: "On the Internet, Nobody Knows You're a Dog,"* 23 COMPUTERS HUM. BEHAV. 3038 (2007). Another summary of the mixed evidence can be found in Chris Baraniuk, *End of Anonymity*, 220 NEW SCIENTIST 34 (2013).

210 *See* Arthur D. Santana, *Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards*, 8 JOURNALISM PRAC. 18 (2014).

real names helped aggressive users gain recognition, increase their "persuasive power," and mobilize followers.[211] This seems consistent with the rise of prominent media personalities who engage in online provocations under their "true" identities, including Alex Jones, Richard Spencer, and Milo Yiannopoulos.[212]

The current climate on Facebook, Twitter, YouTube, and other social media sites is full of bad behavior, even while these companies employ various forms of authenticity enforcement.[213] As danah boyd observed in 2012, "both Facebook and face-to-face settings continue to be rife with meanness and cruelty."[214] Over time, claims that authenticity is a value which encourages people to behave responsibly online—and that we must all "be authentic" so we can collectively reap the benefits of a more civil internet—have failed to match our real-world experiences with social media.

4. Authenticity and Crime

Social media companies rarely discuss authenticity enforcement as a way to get criminals off their networks. This may be because companies are reluctant to remind users that their networks are full of criminals. Nonetheless, authenticity enforcement likely helps reduce fraud, foreign election interference, and other crimes. Facebook has sued some of its own users under the Computer Fraud and Abuse Act ("CFAA") for operating scams that steal users' data.[215] In fact, some aspects of the CFAA have likely encouraged companies to track users' identities.[216]

---

211 *Id.*

212 *See, e.g.*, Ben Schreckinger, *The Alt-Right Comes to Washington*, POLITICO (Jan.–Feb. 2017), https://www.politico.com/magazine/story/2017/01/alt-right-trump-washington-dc-power-milo-214629 (describing the presence of Alt-Right personalities in online chatrooms, Twitter, and spaces on the Internet).

213 *See* Bill Reader, *Free Press vs. Free Speech? The Rhetoric of "Civility" in Regard to Anonymous Online Comments*, 89 JOURNALISM & MASS COMM. Q. 495, 506 (2012) (observing that "the undesirable impoliteness and rudeness found in many online forums appears to accurately reflect the state of the culture, or at least the dominant voices in the culture").

214 danah boyd, *The Politics of "Real Names,"* 55 COMMS. ACM 29, 30 (2012).

215 *See* Counterfeit Access Device and Computer Fraud and Abuse Act of 1984 (CFAA), Pub. L. No. 98-473, § 2102(a), 98 Stat. 2190, 2190–92 (1984).

216 The law prohibits "unauthorized access" to a computer. *Id.*; *see* Thomas E. Kadri, *Digital Gatekeepers*, 99 TEXAS L. REV. (forthcoming 2021) (critiquing cyber-trespass laws like the CFAA that give online platforms gatekeeper rights to block external research); Thomas E. Kadri, *Platforms as Blackacres*, 68 UCLA L. REV. (forthcoming 2021) (outlining possible First Amendment challenges to using cyber-trespass laws to shield online platforms from external scrutiny). Years ago, companies may have developed identity-policing tools in part to preserve their legal options under the CFAA. The use of a false identity in violation of a platform's terms of service might establish unauthorized access,

The underlying concern, however, is criminal conduct, not identity violations. Some criminals hijack strangers' accounts, falsely assuming their targets' identities in the process. These people present "inauthentic identity," but they are *bad actors* in an intuitively criminal sense: they have stolen someone's identity. In other cases, criminals create fictitious identities for the purpose of engaging in crimes like fraud. What is wrong about this behavior is not that the individual is claiming a false identity, but rather that they are committing crimes.

On the other end of the spectrum, as previously noted, some people, including transgender individuals and human rights activists, choose identities that violate authenticity rules as a form of self-expression, or to protect themselves from harm. Others, like low-income people and undocumented immigrants, have "inauthenticity" foisted upon them. These individuals are not committing crimes, but they pay a price when they are caught in violation of authenticity requirements.

5. Collaborating with the State, Forestalling Regulation

Authenticity policies give companies significant power to connect the identities of speakers on their networks to real-world identities—a useful power for collaborations between companies and law enforcement. By exploiting their ability to provide law enforcement with specific, identifying information about targeted individuals, companies may shore up support from public institutions or the public, and reduce calls for regulation.[217] Collaboration between private companies and law enforcement is explored in more length in the next subsection.

* * *

and some scholars have taken the position that unauthorized access can be established if a user returns to a service after having been kicked off. *See, e.g.*, Orin Kerr, *More Thoughts on the Six CFAA Scenarios About Authorized Access vs. Unauthorized Access*, VOLOKH CONSPIRACY (Jan. 28, 2013), http://volokh.com/2013/01/28/more-thoughts-on-the-six-cfaa-scenarios-about-authorized-access-vs-unauthorized-access/ (positing that "future accesses" are unauthorized under the CFAA where a person was previously banned). Where liability is only created upon the creation of a follow-on (presumably false) account or identity, authenticity regulation may be essential to help companies establish the repeat offense and identify wrongdoers to pursue. In other words, authenticity regulation may help companies protect their legal rights in ways that content-based regulation does not.

217 *See, e.g.*, Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. REV. 27, 32 (2019) ("[A]s the Internet grew and became commercialized, platforms became increasingly susceptible to government control and pressure to extend the reach of local law.").

This Section has examined the industry's own claims about why authenticity matters to speech, and it finds that some of them have merit. In particular, *in*authenticity is a speech problem in two main situations: when a speaker has lied about an identity attribute that bears on the truthfulness of her speech, and when a speaker has lied about an identity attribute to participate in an expressive community from which she would otherwise be excluded. In these situations, authenticity regulation would produce benefits to others (not to the speaker herself). Also, in general, people may feel that a network is more trustworthy when people operate under "real" identities, and that a network is less trustworthy when there are many pseudonymous accounts. Certainly, the social media industry, by treating authenticity as an important value, encourages users to hold these beliefs. But is authenticity a moral value? No. Is it effective at reducing abuse, discrimination, and harassment online? Perhaps not. Does it help identify criminals? Some, but it is both under- and over-inclusive.

B. Has Authenticity Been Oversold?

1. The Value of Anonymous and Pseudonymous Speech

Authenticity regulation prohibits "false identity" alongside anonymous and pseudonymous identity (although some companies, like Twitter, allow pseudonymous accounts).[218] And as we have seen, companies sometimes justify this approach on the ground that "authentic" speakers produce "authentic content," which implies that content produced by authentic speakers is truthful and good.[219]

The industry's rejection of anonymous and pseudonymous speech represents a significant break from norms of American political

---

218 Of course, "false" identity is not the same thing as "anonymous" or "pseudonymous" identity. False identity conveys the understanding that the speaker has engaged in deception, claiming attributes that do not truthfully apply.

219 Facebook's Community Standards now state that authenticity is an important value for the company because Facebook "want[s] to make sure the content people are seeing on Facebook is authentic." *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/ (last visited Apr. 1, 2020). The sentence conflates authentic content with authentic identity, implying that a person who presents an authentic identity produces authentic content. This is a semantic trick, because many people would understand "authentic content" to mean that the substance of the content is genuine, true, or accurate—not just that it comes from a speaker operating under his or her real identity.

discourse[220]—and free speech jurisprudence[221]—which have long recognized value in anonymous and pseudonymous speech.[222] A recent example is the *New York Times*' 2018 publication of an anonymous op-ed, authored by an unidentified member of the Trump Administration, which shed light on the internal workings of the presidential administration.[223] The fact that Americans sometimes find anonymous and pseudonymous political expression to be valuable and trustworthy—and that Americans have centuries of experience at evaluating the credibility of anonymous and pseudonymous speech—highlights how private authenticity regulation is challenging longstanding free speech values.[224]

Two traditions have expressed the purpose of free speech in First Amendment jurisprudence: a liberal tradition, which emphasizes the individual's right to expressive liberty, and a republican tradition, which

---

220 *See* Alfred York, *Anonymity, Pseudonymity, and Deliberation: Why Not Everything Should Be Connected*, 26 J. POL. PHIL. 169, 172 (2018) ("Writing under an assumed name or no name at all has long been practiced in domains ranging from literature to philosophy to political argument; indeed, the set of essays published under the pseudonym 'Publius' count among the most notable contributions to American political thought and underpinned public debate on the ratification of the United States Constitution.").

221 *See, e.g.*, Lyrissa Barnett Lidsky & Thomas F. Cotter, *Authorship, Audiences, and Anonymous Speech*, 82 NOTRE DAME L. REV. 1537, 1538 (2007) ("[T]he First Amendment, as interpreted by the United States Supreme Court, confers upon authors a right to speak anonymously or pseudonymously, even when doing so interferes with audiences' attempts to decode their messages."); MacIntyre v. Ohio Elections Comm'n, 514 U.S. 334 (1995) (holding that a statute that prohibited anonymous political or campaign literature unconstitutional).

222 There is a significant literature on anonymity and speech. Unfortunately, a rich and detailed treatment of the subject is outside the scope of this Article. This Subsection focuses on the value of anonymity and pseudonymity for political discourse, but anonymity is well-recognized for its importance to artistic expression and the expression of self (i.e., authenticity in the social psychology sense). *See, e.g.*, Edward Stein, *Queers Anonymous: Lesbians, Gay Men, Free Speech, and Cyberspace*, 38 HARV. C.R.-C.L. L. REV. 159, 163 (2003) ("[R]estrictions on anonymity uniquely affect 'closeted' lesbians, gay men, and other sexual minorities.").

223 Opinion, *I Am Part of the Resistance Inside the Trump Administration*, N.Y. TIMES (Sept. 5, 2018), https://www.nytimes.com/2018/09/05/opinion/trump-white-house-anonymous-resistance.html.

224 For a summary of the benefits and drawbacks of anonymity and pseudonymity in discourse, see Lyrissa Barnett Lidsky & Thomas F. Cotter, *Authorship, Audiences, and Anonymous Speech*, 82 NOTRE DAME L. REV. 1537, 1559–77 (2007). After thoroughly summarizing the pros and cons, the authors argue that, in public law, assuming more speech is better than less speech, and that listeners are "largely rational and capable of self-governance," a "constitutional privilege" in favor of anonymous and pseudonymous speech is preferable to a presumption against it. *Id.* at 1577, 1589–90. In online discourse, both assumptions are less clearly true; the volume of online posts and tweets is vast (so more speech may not be better than less), and the presentation of information on social media may make it particularly difficult for listeners to discern signals of reliability.

promotes public values and self-government.[225] Speaker anonymity or pseudonymity is not a problem in the liberal tradition, since the speaker retains the power to express her identity as she sees fit. In fact, speaker anonymity or pseudonymity likely *enhances* expressive liberty, since some individuals will feel freer to express themselves under an assumed identity.[226]

Speaker anonymity or pseudonymity *does* create potential problems under the republican tradition, however, insofar as a link exists between anonymity or pseudonymity and false and misleading content, which would undermine collective self-government. But people operating under their "true" identities spread false and misleading content all the time.[227] This highlights the misfit between authenticity enforcement and the potential harms of inauthenticity which are offered to justify it. If the core problems are crime, abusive speech, and false and misleading content, solutions should be tailored to fit those problems. Conventional speech norms also trust listeners to assess anonymous and pseudonymous speech as they see fit, rather than encourage them to defer to a third-party decider.[228]

2. Is All Authentic Speech of Equal Worth?

Authenticity regulation gives a green light to speakers who are willing to "own" offensive or false content, and treats the content of the speech of all "authentic" speakers as having roughly equal value (i.e., equally deserving of

---

225 *See* Morgan N. Weiland, *Expanding the Periphery and Threatening the Core: The Ascendant Libertarian Speech Tradition*, 69 STAN. L. REV. 1389, 1404–09 (2017) (examining the two traditions of speech jurisprudence).

226 On the other hand, anonymity and pseudonymity might chill speech if it enables trolling and harassment that silences them. But, as we saw, *supra* Part III.A, the evidence on this is mixed.

227 Speakers operating under their "real" identities routinely circulate misinformation on social media. In 2019, for example, Twitter's Chief Executive Officer, Jack Dorsey, caused controversy by tweeting out praise of Ben Greenfield, a prominent (and verified) health podcaster who is known for his anti-vaccine tweets. *See, e.g.*, Julia Alexander, *Jack Dorsey's Endorsement of Anti-Vax Podcaster Highlights Twitter's Misinformation Problem*, VERGE (Mar. 13, 2019), https://www.theverge.com/2019/3/13/18264196/jack-dorsey-anti-vax-ben-greenfield-twitter-facebook-youtube-amazon-conspiracy (reporting the controversy surrounding Jack Dorsey's tweet); Ben Greenfield @bengreenfield, TWITTER (Feb. 11, 2019, 7:49 AM), https://twitter.com/bengreenfield/status/1094986690785988613? ("Vaccines do indeed cause autism . . . ."); *see also* Alexandria Neason, *On Twitter, News Outlets Amplify Trump's False Statements: Study*, COLUM. JOURNALISM REV. (May 3, 2019), https://www.cjr.org/politics/twitter-media-trump.php (discussing the spread of false and misleading statements made on Twitter by President Donald J. Trump).

228 *See* Lidsky & Cotter, *supra* note 224, at 1539 (noting that "audiences are likely to discount the value of nonattributed speech, thus mitigating some (but not all) of anonymous speech's potential harm").

protection from censorship).[229] By not addressing the content of the offensive speech directly, companies fail to communicate that some authentic speech is false, misleading, degrading, or abusive.

Fundamentally, authenticity regulation teaches that it's not the content of speech that is objectionable, it's *the person who is doing the speaking*. Speech communicated by one actor might be a violation—while the very same content, communicated by a different actor, is perfectly fine. For example, on Facebook, only an inauthentic speaker who says dehumanizing things about homeless children is doing anything wrong.[230] An authentic speaker communicating the same content is not violating Facebook's rules, and is not treated as blameworthy. This approach to objectionable speech is quite different from the approach in First Amendment law, which has generally acknowledged that some speech is both objectionable *and* protected from censorship. In addition, the industry's take on objectionable speech—hate the speaker, not the speech—gets the merits exactly backwards. The substance of the speech—in the example offered above, dehumanizing homeless children—*is* the problem. The person *expressing* the hateful content may be capable of rehabilitation.

The problem goes beyond "normalization." By choosing to tolerate noxious speech produced by "authentic" speakers, companies permit such speakers to leverage the companies' powerful communicative technology. This gives those speakers the ability to integrate their ideas into the industry's machine learning; to pay to "push" their ideas into others' news feeds; and to employ identity-based targeting and exclusions to maximize the persuasive effect of their speech (and their advertising dollars). For example, Facebook's choice to provide a platform for racists who are willing to "own" their racism may indeed make racism seem normal and acceptable. However, it also provides a channel to deliver racist ideas in a way designed to achieve maximum persuasion, and to inject racist expression into Facebook's machine intelligence, where it affects pattern recognition and influences future customization and expression.

---

229 *See Terms of Service*, FACEBOOK, https://www.facebook.com/terms.php?ref=p (last visited Apr. 1, 2020) ("When people stand behind their opinions and actions, our community is safer and more accountable.").

230 Facebook's content rules prohibit dehumanizing speech against only some groups. These do not include either children or the homeless. *See Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards/objectionable_content (last visited Apr. 1, 2020).

3. Inauthentic Behavior

Part I documented Facebook's original practice of defining authenticity in terms of a person's "real" or "true" identity, and its recent move to redefine authenticity in terms of behavior. Under the new approach, a person can present her "true" identity on Facebook and yet still run afoul of its authenticity rules. Coordinated inauthentic behavior is "when multiple accounts—including both fake and authentic accounts—work together to mislead people."[231] Deceitful *behavior* is what puts a user in violation of Facebook's rules against coordinated inauthentic behavior. Facebook has been very clear that it is "taking down these Pages and accounts based on their behavior, not the content they post."[232]

The factors that Facebook considers relevant to evaluating the "authenticity" of behavior mainly relate to speech and association. As Facebook's General Counsel testified in October 2017, "[o]ur systems examine thousands of account attributes and focus on detecting behaviors that are very difficult for bad actors to fake, including their connections to others on our platform."[233] Facebook has said that it "can find links between accounts that might be coordinating an information operation based on how they interact on Facebook or other technical signals that link the accounts together."[234] Essentially, the company looks for points of association between a suspected bad actor and other accounts.[235] In October 2018, for example,

---

231 Testimony of Sheryl Sandberg, *supra* note 48, at 3. Sandberg explained that coordinated inauthentic behavior "is not allowed because we don't want organizations or individuals creating networks of accounts that misinform people about who they are or what they're doing." *Id.*

232 Nathaniel Gleicher, *Removing Coordinated Inauthentic Behavior from Russia*, FACEBOOK NEWSROOM (Jan. 17, 2019), https://newsroom.fb.com/news/2019/01/removing-cib-from-russia/.

233 Testimony of Colin Stretch, *supra* note 2, at 3.

234 Nathaniel Gleicher, *More Information about Last Week's Takedowns*, FACEBOOK NEWSROOM (Nov. 13, 2018), https://newsroom.fb.com/news/2018/11/last-weeks-takedowns/.

235 The Digital Forensic Research Lab, which has studied coordinated inauthentic behavior, has focused on the "pattern of connections" between accounts and pages in evaluating authenticity. Dig. Forensic Res. Lab, *Facebook's Sputnik Takedown—In Depth*, MEDIUM (Jan. 17, 2019), https://medium.com/dfrlab/facebooks-sputnik-takedown-in-depth-f417bed5b2f8. "Cross-posting" is an example. In January 2019, Facebook removed accounts and pages on the basis of coordinated inauthentic behavior. Among the behaviors that implicated the accounts and pages in coordinated inauthentic behavior was cross-posting of videos. *Id.* The Digital Forensic Research Lab has written that cross-posting "proves that there is a relationship between two pages," which serves as a basis for finding coordinated inauthentic behavior. *Id.* (On Facebook, accounts "can only cross-post one another's content if both agree to it or if they already have an administrator or manager in common" and "@DFRLab identified different patterns of cross-posting and sharing videos between" Pages implicated in the coordinate inauthentic behavior). Different Pages that upload the same videos separately also raise behavioral flags. *Id.* And Facebook has pointed to

after Facebook banned Gavin McInnes, the founder of a reputed hate group, for violations of its hate speech policies, it proceeded to remove "both individual accounts and pages, as well as associated groups, that [were] affiliated" with the group online.[236] Facebook has also charged users with employing coordinated inauthentic behavior "where a Page name was changed after it had built up a large following, substantially changing the Page's subject matter."[237]

In August 2018, Facebook removed a number of accounts and Pages, citing coordinated inauthentic behavior, including Pages run by American anti-racism activists.[238] Facebook said that it had "observed links" between Russian propaganda groups and a group that created a Facebook event Page for an anti-racism rally in Washington D.C.[239] The event, however, was real. Smash Racism, a grass-roots organization that co-sponsored the rally, issued a statement that said, in part:

> Facebook's removal of the page in question is censorship against the real movement against white supremacy and fascism. The only evidence connecting our page to Russia/the Internet Research Institute is a single admin account for the Resisters, which was an admin for 7 minutes. All other evidence (such as use of VPNs and sock accounts) represent common practices for anti-fascists in today's climate.[240]

What this suggests, of course, is guilt by association: if one user takes steps to amplify content posted by a bad actor, that person has become part of the

---

similar posts being shared by different Pages "in a coordinated way" as evidence of inauthentic behavior. Nathaniel Gleicher, *Banning Twinmark Media Enterprises in the Philippines from Facebook*, FACEBOOK NEWSROOM (Jan. 10, 2019), https://newsroom.fb.com/news/2019/01/banning-twinmark-media-enterprises/. When different accounts work together to amplify content in these and other ways, they are flagged by Facebook as "networks of accounts" attempting to "mislead others about who they were and what they were doing." Gleicher, *supra* note 232.

236 Nick Statt, *Facebook Bans Accounts Affiliated With Far-Right Group the Proud Boys and Founder Gavin McInnes*, VERGE, (Oct. 30, 2018, 8:27 PM), https://www.theverge.com/2018/10/30/18045410/facebook-bans-proud-boys-far-right-extremist-group-gavin-mcinnes (quoting a Facebook spokesperson as saying that Facebook not only bans hate groups and associated individuals, but also "remove[s] all praise and support when we become aware of it"). The company did *not* cite "coordinated inauthentic behavior" as the basis for the takedowns, but rather "violations of its rules on hate speech and the organizing of groups that spread hate both online and offline." *Id.*

237 Gleicher, *Banning Twinmark Media Enterprises*, *supra* note 235.

238 *See* Elias Groll, *Anti-Racism Groups Feel Tarred by Facebook's Fight Against Fake Accounts*, FOREIGN POLICY (Aug. 1, 2018, 8:03 PM), https://foreignpolicy.com/2018/08/01/anti-racism-activists-furious-facebook-smears-protest-with-russian-link/ (analyzing why Facebook deleted an account over an event that turned out to be real).

239 *Id.*

240 Smash Racism DC, *A Statement from the Shut It Down Coalition on Facebook's false "Russian Bot" censorship*, FACEBOOK NOTES (July 31, 2018), https://www.facebook.com/notes/smash-racism-dc/a-statement-from-the-shut-it-down-dc-coalition-on-facebooks-false-russian-bot-ce/1310411682428767/.

"network." The First Amendment protects the right of association[241]; Facebook treats the wrong kinds of associations as evidence of prohibited inauthenticity.

The shift to treating "inauthentic identity" as a behavior also signals a particular view about political persuasion. In January 2019, Benjamin Decker, a research fellow at the Shorenstein Center on Media, Politics and Public Policy, gave an interview to the *Mercury News* about Facebook's removal of the accounts of five left-leaning technology experts. Facebook had banned them for engaging in "coordinated inauthentic behavior" by creating a Facebook page with a conservative name. Decker told journalists that "it was inauthentic and misleading for a left-leaning political operative to try to create communities of conservatives for the express purpose of sending those people political messages that would sway their thinking—and to use the label of a news organization to do so."[242] Of course it is misleading to claim to be a news organization if you're not, but is it misleading to create a web page to appeal to your opponents in order to "sway their thinking," if you acknowledge your real identity?

4. Authenticity as Attack Strategy

Authenticity regulation has evolved into an effective means for one party to attack an opponent. Most social media companies rely heavily on user reports of rule violations, including reports of "fake accounts." When a company receives a report that a user is publishing under a false identity, it is common for the company to demand that the user verify his or her identity. Unlike content-based attacks, which also occur but are limited to pieces of content, authenticity-based attacks are particularly potent. They can result in the temporary suspension of a whole account until verification requirements are satisfied. This sort of offensive strategy burdens the speech of the victim, even if he or she is operating under a "true" identity and is eventually exonerated. Reports suggest that this kind of abusive tactic is common.[243]

---

241 *See, e.g.*, NAACP v. Claiborne Hardware Co., 458 U.S. 886, 918–19 (1982) (holding that the First Amendment restricts the ability of the State to impose liability on an individual solely because of their association with another actor).

242 Tony Romm et al., *Facebook Investigates Group Backed By Reid Hoffman*, MERCURY NEWS (Jan. 8, 2019, 6:50 AM), https://www.mercurynews.com/2019/01/08/facebook-investigates-group-backed-by-reid-hoffman/ (paraphrasing statements made by Decker).

243 *See, e.g.*, Brett Solomon, *What Can Social Media Platforms Do For Human Rights?*, OPENDEMOCRACY (Oct. 26, 2015), https://www.opendemocracy.net/en/what-can-social-media-platforms-do-for-human-rights/ ("For years, people have been harassed on Facebook by adversaries who flag them as having 'fake' identities, even when they're using their real names.").

5. Micro-Targeting and Discourse

Micro-targeting itself, which is controlled and implemented with little outside scrutiny by private companies, raises alarms. Although it seems reasonable for speakers to be able to direct their speech at a particular audience in a public way—for example, by taking out an advertisement in the *Wall Street Journal* or by running a commercial on a cable television network like the Food Network—it also seems *un*reasonable for social media companies to earn profits by charging fees to exclude short people, or people with diabetes, or men, from targeted political advertising. Preliminary research has shown that Americans disapprove of advertisement targeting, and that individuals with lower incomes and lower educational attainment levels are less likely to notice it.[244] So, micro-targeting can shape political discourse without recipients realizing that they are receiving very different speech from others. And because some elements of micro-targeting involve black-box proprietary algorithms—not just choices exercised by speakers—micro-targeting can provide companies with opportunities to engage in viewpoint discrimination. It is not clear how this discrimination would ever become apparent to users or the public.

Political advertisement transparency initiatives implemented to date—such as the advertisement archive established by Facebook—do not provide information about micro-targeting, and as a result we lack good information about how social media companies and their clients are using micro-targeting to shape debate.

6. The "Right and Privilege" to Evaluate Speech on Its Own Merit

Citizens United v. FEC[245] articulated the Court's boldest-ever arguments opposing speaker-based discrimination and subsequent cases have continued to develop these themes.[246] The Supreme Court reasoned that speaker-based

244 *See* Russell Heimlich, *Internet Users Don't Like Targeted Ads*, PEW RES. CTR. (Mar. 13, 2012), https://www.pewresearch.org/fact-tank/2012/03/13/internet-users-dont-like-targeted-ads/ (finding a majority of every demographic group dislike online targeted advertising).

245 558 U.S. 310 (2010).

246 *See* Nat'l Inst. of Family & Life Advocates v. Becerra, 138 S. Ct. 2361, 2375–78, (2018) (finding that a disclosure law targeting licensed and unlicensed crisis pregnancy clinics was speaker-based); Sorrell v. IMS Health Inc., 564 U.S. 552, 563–64 (2011) (holding that a Vermont law that engaged in content- and speaker-based discrimination violated the First Amendment); Ariz. Free Enter. Club's Freedom Club PAC v. Bennett, 564 U.S. 721, 727–28 (2011) (holding that a "matching funds scheme" substantially burdened political speech and therefore violated the First Amendment); Minn. Citizens Concerned for Life, Inc. v. Swanson, 692 F.3d 864, 871 (8th Cir.

discrimination harms both speakers and listeners. It harms speakers by taking the right to speak from some and giving it to others, infringing the speaker's "right to use speech" to "strive to establish worth, standing, and respect for [its] voice."[247] It harms listeners because "the public" has "the right and privilege" to evaluate speech on its merits.[248] "This Court's precedents are deeply skeptical of laws that 'distinguis[h] among different speakers, allowing speech by some but not others,'" the Court explained in a 2018 case.[249]

Authenticity rules prevent social media users from evaluating speech from "inauthentic" speakers on its merits. Thus, during a period in which the Supreme Court has increased its hostility to speaker-based discrimination, private ordering by social media companies has moved in the opposite direction, evolving *in favor* of speaker-based strategies. To the extent that the Supreme Court's opposition to speaker-based discrimination is primarily grounded in suspicions about viewpoint discrimination or manipulation of public debate, these concerns extend to social media companies.[250]

7. Identity Theft and the State Apparatus

The social media industry's reliance on authenticity regulation has led to two related developments: an arms race between the industry and identity thieves, and a strong "mutuality of interest" between the industry and law

---

2012) (discussing the Supreme Court's disapproval of speaker-based discrimination in the context of political speech); *see also* Dallman v. Ritter, 225 P.3d 610, 634 (Colo. 2010) (holding that a Colorado constitutional amendment prohibiting campaign donations from organizations receiving single-source government contracts did not sufficiently serve an important government interest and therefore, violated the First Amendment).

247 *Citizens United*, 558 U.S. at 340–41.

248 *Id.* Although the case concerned a speech ban, this part of the opinion bridged a connection to the argument, articulated in earlier Supreme Court opinions, that a speaker's identity is irrelevant to an evaluation of his or her speech. *See, e.g.*, First Nat'l Bank v. Bellotti, 435 U.S. 765, 777 (1978) ("The inherent worth of the speech in terms of its capacity for informing the public does not depend upon the identity of its source, whether corporation, association, union, or individual."); *see also* Doe v. Reed, 561 U.S. 186, 238–39 (2010) (Thomas, J., dissenting) (applying *Bellotti* to referendum measures).

249 *Becerra*, 138 S. Ct. at 2378 (adding that "speaker-based laws run the risk that 'the State has left unburdened those speakers whose messages are in accord with its own views.'") (quoting *Sorrell*, 564 U.S. at 580).

250 Michael Kagan, *Speaker Discrimination: The Next Frontier of Free Speech*, 42 FLA. ST. U. L. REV. 765, 767 (2015) ("*Citizens United* should be understood as articulating and explaining a set of principles that have long been implicit in the case law.").

enforcement.[251] The industry's choice to police speech through authenticity means that a user can get around its rules by stealing the identity of a real person. That is, the skilled identity thief can avoid authenticity violations, which are not crimes, with real crime. Since some identity thieves (say, Russian intelligence) are likely to be more technologically sophisticated than others (say, ordinary teenagers), the speakers who cause the most harm might not be the speakers who are easiest to catch. Under a system of authenticity regulation, we might expect identity theft generally to increase and to grow more sophisticated, particularly in the lead-up to elections. Notice, also, that as Facebook increases incentives for identity thieves to steal identities, the company makes its own identity-verification capabilities more valuable.[252]

One problem with the authenticity approach is that it potentially locks companies into an arms-race with foreign state powers, which have a head-start, vast resources, and technological prowess to evade detection. Russia, like other foreign nations, has a significant intelligence apparatus developed over many decades. Much spycraft is specifically geared toward hiding the identities of agents. To fight back, companies have formed logical—perhaps *necessary*—alliances: they have joined forces with the U.S. Government. Through this public-private partnership, Facebook and other companies fight the "inauthenticity" of foreign spies, shoulder-to-shoulder with the Trump Administration.

In August 2018, a sharp reporter asked Facebook executives how the company identified "bad actors" to remove from its platform. It was an important question: How does Facebook determine that a user's identity is "inauthentic," justifying his or her removal?

Facebook's response suggested that the company has relied heavily on the U.S. Government to identify "bad actors." In reply, a Facebook executive seamlessly adopted the jargon of the U.S. intelligence community. "[T]hese assets," he said, referring to the "bad actors," "have been previously identified—not necessarily by us, but by intelligence services in the U.S.—as

---

251 ZUBOFF, *supra* note 10, at 19 (noting "mutuality of interests between fledgling surveillance capitalists and state intelligence agencies").

252 It may be easier for identity thieves (e.g., Russian state-sponsored identity thieves) to steal the identities of Americans who are not Facebook users, and create new Facebook accounts for them. Facebook will likely have a harder time identifying that the person's identity has been stolen, because Facebook has a lot less identifying information about non-users and because a non-user is less likely to notice that someone is posting on Facebook using his or her identity.

linked to Russian intelligence agencies."[253] In other words, in at least some cases, Facebook is taking the Government's word for it.[254] In 2019, Facebook said that it is "constantly" working to "stay ahead" of "bad actors" by "building better technology, hiring more people and working more closely with law enforcement, security experts and other companies" whose "collaboration was critical to [recent] investigations."[255]

Collaboration between the State and powerful social media companies raises censorship concerns and strengthens arguments that First Amendment law should treat Facebook and its competitors as "state actors."[256] Certainly, when facing off against Russian foreign intelligence services, it is helpful to have the U.S. intelligence community as an ally. However, there is real danger that the Federal Government will use companies to suppress speech it does not like by labeling disfavored speakers as bad or inauthentic actors.[257] This is particularly effective where companies' identity-verification systems

253 *Press Call Transcript*, FACEBOOK NEWSROOM (Aug. 21, 2018, 4:30 PM PST), https://fbnewsroomus.files.wordpress.com/2018/08/8-21-press-call-transcript.pdf (statement of Nathaniel Gleicher).

254 *See also* Gleicher, *supra* note 232 (stating that a January 2019 takedown of 107 Facebook Pages, Groups, and accounts, and forty-one Instagram accounts, was the result of "an initial tip from US law enforcement").

255 *Id.*; *see also* Gleicher, *supra* note 234 ("On November 4 [2018], the FBI tipped us off about online activity that they believed was linked to foreign entities. Based on this tip, we quickly identified a set of accounts that appeared to be engaged in coordinated inauthentic behavior . . . . "). Another post further justified the company's reliance on the State to identify targets for censorship, noting that "law enforcement agencies can draw connections off our platform to a degree that we simply can't." Nathaniel Gleicher, *How We Work with Our Partners to Combat Information Operations*, FACEBOOK NEWSROOM (Nov. 13, 2018), https://newsroom.fb.com/news/2018/11/last-weeks-takedowns/. The author explained that "[t]ips from government and law enforcement partners can therefore help our security teams attribute suspicious behavior to certain groups, make connections between actors, or proactively monitor for activity targeting people on Facebook." *Id.*

256 Major social media competitors, like Facebook and Twitter, have long admitted that they share information and work in unison to silence inauthentic speakers. *See, e.g.*, Testimony of Colin Stretch, *supra* note 2, at 3. (stating that Facebook reaches out "to leaders in our industry and governments around the world to share information on bad actors and threats so that we can make sure they stay off all platforms"); Gleicher, *How We Work With Our Partners to Combat Information Operations*, *supra* note 255 ("[W]e've worked closely with our fellow tech companies, both bilaterally and as a collective, to deal with the threats we have all seen during and beyond elections."); Tony Romm & Craig Timberg, *Facebook Suspends 'Inauthentic' Iranian Accounts that Criticized Trump and Spread Divisive Political Messages*, WASH. POST (Oct. 26, 2018), https://www.washingtonpost.com/technology/2018/10/26/facebook-suspends-inauthentic-iranian-accounts-that-criticized-trump-spread-divisive-political-messages/ ("Twitter said it had removed a small number of accounts based on information Facebook supplied.").

257 *See, e.g.*, *Press Call Transcript*, *supra* note 253 (statement of Nathaniel Gleicher) (explaining, in answer to a reporter's question, that Facebook had removed Pages of "bad actors" because "these assets have been previously identified—not necessarily by us, but by intelligence services in the U.S.—as linked to Russian intelligence agencies.").

impose prior restraints on speech. The possibility that authenticity regulation could be utilized by the Federal Government to suppress Americans' speech should cause us to ask hard questions about its methods.[258]

8. Commodifying Identity

By focusing regulatory enforcement on identity rather than content, the social media industry has turned an authentic digital identity into a valuable asset. Already, it is a common practice for users to sell administrative rights to existing Pages on Facebook, an act that the company prohibits.[259] In addition, a person can pay a *proxy* to communicate the person's objectionable speech, using the proxy's own authentic (or verified) identity. Verification services and scams have proliferated online.[260]

By commodifying approved identities, the social media industry has not solved problems caused by unapproved identities; it has merely created offline markets to exploit the authenticity rule systems. One danger is that this will replicate the kinds of identity-nesting that have posed problems for years in other areas, such as tax evasion and campaign finance. Another is that it will simply advantage wealthy speakers, who can leverage their resources to exploit loopholes in the system.

* * *

In the final analysis, the benefits of authenticity regulation for speech seem outweighed by the dangers they present to a range of important interests. A main danger of any system of speech regulation, public or private, is that it grants the regulator unfettered power to silence viewpoints with which it disagrees. As the foregoing makes clear, authenticity regulation creates opportunities for viewpoint discrimination, just like content

---

258 *See, e.g.*, Michele Gilman & Rebecca Green, *The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization*, 42 N.Y.U. REV. L. & SOC. CHANGE 253, 257 (2018) ("Increasingly, large-scale data sharing between different levels of government and private industry blurs public/private distinctions.").

259 *See, e.g.*, Paige Occeñola & Geno Gonzales, *PH Company Banned By Facebook Spread Lies, Used Fake Accounts*, RAPPLER (Jan. 17, 2019, 1:57 PM), https://www.rappler.com/technology/social-media/220741-facebook-remove-trending-news-portal-twinmark-media-enterprises (reporting that Twinmark Media Enterprises, an organization banned by Facebook in January 2019, "was selling admin rights to Facebook Pages it had created, in order to increase distribution and generate profit").

260 *See* Taylor Lorenz, *The Problem with Verification*, ATLANTIC (June 25, 2019), https://www.theatlantic.com/technology/archive/2019/06/instagram-and-twitter-should-eliminate-verification/592351/ ("Hundreds of people online advertise verification services. And some users have even been able to obtain a check mark after paying thousands of dollars.").

moderation does. It also invites law enforcement to participate in flagging inauthentic speakers. However, unlike content moderation, authenticity regulation may obscure viewpoint discrimination when it happens. What the public sees is a clever way to monetize data, or even a righteous purge of "bad actors." This is a danger that deserves serious consideration.

CONCLUSION

This Article has drawn close connections between authenticity regulation in the social media industry and the industry's business imperatives. Authenticity rules provide social media companies with significant business value. They facilitate companies' analytics-based business models. And, increasingly, they tap into a new speech ethos—which the companies themselves are largely responsible for popularizing—which treats authenticity as a moral virtue; as a behavior that can be policed; as a proxy for "authentic content"; and as a value analogous to human rights like privacy and dignity. Authenticity, in the industry sense, has multi-dimensional, evolving meanings.

The use of authenticity regulation by social media companies deserves greater attention from the legal academy—not only its potential to incorporate bias, and its broader implications for identity, dignity, expression, and democratic discourse, but also its capacity to suppress viewpoints and shape information flows. Today, Facebook removes more speech from its network for violations of its authenticity rules than for violations of its content-based rules, but with considerably less critical scrutiny by journalists and scholars, and with less transparency and oversight. Companies' authenticity enforcement decisions are shared with other firms and become de facto industry-wide takedowns, and information obtained from users who agree to follow authenticity rules can be opportunistically shared with state actors, such as law enforcement.

Is authenticity, as enforced by the social media industry, an essential speech value? This Article argued that authenticity has *some* value for online speech, but mainly as a stand-in for truthfulness, which companies refuse to regulate. As a (bizarre) result, under most companies' rules, a social media user can lie about any subject but herself.