1-1-2019

Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech

Thomas E. Kadri
Assistant Professor of Law *University of Georgia School of Law*, tek@uga.edu

Kate Klonick
Assistant Professor of Law *St. John's University School of Law*, klonickk@stjohns.edu

FACEBOOK V. *SULLIVAN*: PUBLIC FIGURES AND NEWSWORTHINESS IN ONLINE SPEECH

THOMAS E. KADRI* & KATE KLONICK†

In the United States, there are now two systems to adjudicate disputes about harmful speech. The first is older and more established: the legal system in which judges apply constitutional law to limit tort claims alleging injuries caused by speech. The second is newer and less familiar: the content-moderation system in which platforms like Facebook implement the rules that govern online speech. These platforms are not bound by the First Amendment. But, as it turns out, they rely on many of the tools used by courts to resolve tensions between regulating harmful speech and preserving free expression—particularly the entangled concepts of "public figures" and "newsworthiness."

This Article offers the first empirical analysis of how judges and content moderators have used these two concepts to shape the boundaries of free speech. It first introduces the legal doctrines developed by the "Old Governors," exploring how courts have shaped the constitutional concepts

---

*. Resident Fellow, Yale Law School; Ph.D. Candidate, Yale University.

†. Assistant Professor of Law, St. John's University School of Law; Affiliate Fellow, Yale Law School Information Society Project. Authors listed alphabetically. Portions of this Article build on research and ideas featured in Thomas Kadri's and Kate Klonick's doctoral dissertations and as well as an essay published in the *Emerging Threats* series organized by the Knight First Amendment Institute at Columbia University. *See* Kate Klonick, *Facebook v.* Sullivan, KNIGHT FIRST AMEND. INST. (Oct. 1, 2018), https://knightcolumbia.org/content/facebook-v-sullivan [https://perma.cc/W9R8-7K9B]. The authors are grateful for the friends and colleagues whose insights improved this piece, especially Jack Balkin, Molly Brady, Aaron Caplan, Danielle Citron, Jennifer Daskal, Evelyn Douek, Sarah Haan, Margot Kaminski, Daphne Keller, Jennifer Rothman, Robert Post, Rory Van Loo, Morgan Weiland, and colleagues at the Yale Information Society Project. Additional thanks for the opportunity to present this work and receive excellent feedback at the Cornell Tech Speed Conference, Loyola Law School Los Angeles Faculty Workshop, Washington School of Law Faculty Workshop, University of Colorado Boulder Silicon Flatirons Conference, UCLA Social Media Conference, and Yale Freedom of Expression Conference. A very special thank you to David Pozen, whose excellent, patient editing and thoughtful feedback consolidated this Article's central arguments and encouraged new ones.

of public figures and newsworthiness in the face of tort claims for defamation, invasion of privacy, and intentional infliction of emotional distress. The Article then turns to the "New Governors" and examines how Facebook's content-moderation system channeled elements of the courts' reasoning for imposing First Amendment limits on tort liability.

By exposing the similarities and differences between how the two systems have understood these concepts, this Article offers lessons for both courts and platforms as they confront new challenges posed by online speech. It exposes the pitfalls of using algorithms to identify public figures; explores the diminished utility of setting rules based on voluntary involvement in public debate; and analyzes the dangers of ad hoc and unaccountable newsworthiness determinations. Both courts and platforms must adapt to the new speech ecosystem that companies like Facebook have helped create, particularly the way that viral content has shifted normative intuitions about who deserves harsher rules in disputes about harmful speech, be it in law or content moderation.

Finally, the Article concludes by exploring what this comparison reveals about the structural role platforms play in today's speech ecosystem and how it illuminates new solutions. These platforms act as legislature, executive, judiciary, and press—but without any separation of powers to establish checks and balances. A change to this model is already occurring at one platform: Facebook is creating a new Oversight Board that will hopefully provide due process to users on the platform's speech decisions and transparency about how content-moderation policy is made, including how concepts related to newsworthiness and public figures are applied.

TABLE OF CONTENTS

INTRODUCTION

In the summer of 2017, a group of American neo-Nazis convened for a "Unite the Right" rally in Charlottesville, Virginia. Amid scenes of chaos, James Alex Fields drove his car through a crowd and killed a young counter-protestor, Heather Heyer. The next day, a blog post from the white supremacist website, *The Daily Stormer*, was shared over 65,000 times on Facebook: "Heather Heyer, Woman Killed in Road Rage Incident was a Fat, Childless 32-Year-Old Slut."[1] To some, this post looked like provocative but permissible commentary about someone who was now newsworthy; to others, it seemed like harmful speech that Facebook should remove as a violation of its internal rules.[2] When people complained about the post, Facebook's rulemakers debated: Should it stay up or come down? Ultimately, they hedged. The platform removed every link to the post unless the user included a caption condemning *The Daily Stormer*.[3]

This episode reveals something important about free speech in the digital age: the judiciary is no longer the only actor that adjudicates claims about harmful speech. In the United States, we now have two systems to adjudicate these disputes. The first is older and more established: the legal system in which judges apply constitutional law to limit tort claims alleging injuries caused by harmful speech.[4] The second is newer and less familiar:

---

1. Talia Lavin, *The Neo-Nazis of The Daily Stormer Wander the Digital Wilderness*, NEW YORKER (Jan. 7, 2018), https://www.newyorker.com/tech/annals-of-technology/the-neo-nazis-of-the-daily-stormer-wander-the-digital-wilderness [https://perma.cc/8XL7-UYKV]; *see also* Casey Newton, *Facebook Is Deleting Links to a Viral Attack on a Charlottesville Victim*, THE VERGE (Aug. 14, 2017, 8:30 PM), https://www.theverge.com/2017/8/14/16147126/facebook-delete-viral-post-charlottesville-daily-stormer [https://perma.cc/JLF2-AT2G].

2. *See* Julia Angwin et al., *Have You Experienced Hate Speech on Facebook? We Want to Hear from You.*, PROPUBLICA (Aug. 29, 2017, 10:05 AM), https://www.propublica.org/article/have-you-experienced-hate-speech-on-facebook-we-want-to-hear-from-you [https://perma.cc/A8PF-MC4U].

3. Newton, *supra* note 1.

4. *See, e.g.*, Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 56 (1988) (intentional infliction of emotional distress); Time, Inc. v. Hill, 385 U.S. 374, 381 (1967) (privacy); N.Y. Times Co. v. Sullivan, 376 U.S. 254, 271 (1964) (defamation). On the history and complications of constitutionalizing tort law,

the content-moderation system in which platforms like Facebook implement the rules that govern online speech.[5] These platforms aren't bound by the First Amendment. But, as it turns out, they rely on many of the tools used by courts to resolve tensions between regulating harmful speech and preserving free expression—particularly the entangled concepts of "public figures" and "newsworthiness."

This Article analyzes how judges and content moderators have used these two concepts to shape the boundaries of free speech. By exposing the similarities and differences between how the two systems have understood these concepts, this Article offers lessons for both courts and platforms as they confront new challenges posed by online speech. Finally, the Article briefly explores how this comparison reveals the structural changes that platforms like Facebook should make to address these challenges and bring oversight to their governance of online speech.

Part I introduces the legal doctrines developed by the "Old Governors," exploring how courts have shaped the concepts of public figures and newsworthiness in the face of tort claims for defamation, invasion of privacy, and intentional infliction of emotional distress. Part II turns to the "New Governors" and examines Facebook's content-moderation system. Drawing on internal Facebook documents and several years of exclusive interviews with current and former Facebook employees, this Part reveals for the first time how and why the platform created its own rules related to public figures and newsworthiness. In so doing, it shows that Facebook's rulemaking—consciously or unconsciously—channeled elements of the courts' reasoning when imposing First Amendment limits on tort liability.

---

see Daniel J. Solove & Neil M. Richards, *Rethinking Free Speech and Civil Liability*, 109 COLUM. L. REV. 1650, 1672–84 (2009); Eugene Volokh, *Tort Liability and the Original Meaning of the Freedom of Speech, Press, and Petition*, 96 IOWA L. REV. 249, 251–54 (2010).

5. For foundational work on the development of the idea of private governance by speech platforms, see REBECCA MACKINNON, CONSENT OF THE NETWORKED 149–65 (2012) (analyzing platforms through the lens of "governance" at the "new digital sovereigns" of "Facebookistan" and "Googledom"); see also Anupam Chander, *Facebookistan*, 90 N.C. L. REV. 1807, 1819–22 (2012); Robert Gorwa, *What is Platform Governance?*, 22 INFO. COMM. & SOC'Y 854, 854 (2019); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1616–18 (2018); Thomas Kadri, *How Supreme a Court?*, SLATE (Nov. 19, 2018, 1:59 PM), https://slate.com/technology/2018/11/facebook-zuckerberg-independent-speech-content-appeals-court.html [https://perma.cc/524G-QYLJ] [hereinafter Kadri, *How Supreme a Court?*]; Thomas Kadri, *Speech vs. Speakers*, SLATE (Jan. 18, 2018, 12:56 PM), https://slate.com/technology/2018/01/twitters-new-rules-blur-the-line-between-extremists-speakers-and-their-speech.html [https://perma.cc/JW9B-FFQ6] [hereinafter Kadri, *Speech vs. Speakers*]; Kate Klonick & Thomas Kadri, Opinion, *How to Make Facebook's 'Supreme Court' Work*, N.Y. TIMES (Nov. 17, 2018), https://www.nytimes.com/2018/11/17/opinion/facebook-supreme-court-speech.html [https://perma.cc/DRD4-N8AK].

Guided by this background, Part III exposes the similarities and differences between how the private and public governance systems have understood the concepts of public figures and newsworthiness. Through this comparative analysis, this Part addresses how both courts and platforms should confront new challenges posed by online speech. Judges and platform policymakers must adapt to the new speech ecosystem that companies like Facebook have helped create, particularly the way that virality has shifted normative intuitions about who deserves harsher rules in disputes about harmful speech, be it in constitutional law or content moderation.[6] This Part first exposes the pitfalls of using algorithms to identify public figures, critiquing Facebook's use of online news aggregators and news sources to determine public-figure status on the platform—a mechanism that results in Facebook removing too much benign speech and preserving too much harmful speech. It then explains how the internet has eroded traditional reasons for specially protecting speech about public figures because these reasons rest principally on assumptions that people become public figures by choice and that, as public figures, they have greater access to channels of rebuttal.[7] These assumptions are becoming increasingly outdated in the digital age, given the dynamics of online speech[8] and the ubiquity of means to engage in counter-speech.[9] As a result, this Part reassesses the normative utility of distinguishing between "voluntary" and "involuntary" public figures to judge who deserves harsher rules in courts and on platforms, discussing the alternative use of "sympathy" as a normative barometer. Lastly, this Part assesses the risks posed by platforms creating ad hoc exceptions for newsworthy content. Although there are significant advantages to increasing human review in content moderation, the current

---

6. *See generally* THOMAS I. EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 3–6 (1970) (developing the idea of the First Amendment creating a "system" of free speech).

7. *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 352 (1974); Curtis Publ'g Co. v. Butts, 388 U.S. 130, 154 (1967).

8. *See* DANAH BOYD, IT'S COMPLICATED: THE SOCIAL LIVES OF NETWORKED TEENS 11–12 (2014) (explaining how social media creates new challenges because of the "persistence," "visibility," "spreadability," and "searchability" of content); Kate Klonick, *Re-Shaming the Debate: Social Norms, Shame, and Regulation in an Internet Age*, 75 MD. L. REV. 1029, 1053–54 (2015) (describing how the internet changes social norm enforcement by eliminating frictions of time, geography, personal reputation and cost). *See generally* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014) (describing cyber harassment and exploring ways to combat it); Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296 (2014) (discussing the changing approach to speech regulation).

9. *See* Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L.J. 1805, 1833 (1995) (predicting, even before the advent of social media, that new technologies would "both democratize the information marketplace—make it more accessible to comparatively poor speakers as well as rich ones—and diversify it"); *see also* LAWRENCE LESSIG, CODE VERSION 2.0, at 19 (2006); Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004).

structure of most platforms means that decisions in this area lack transparency, expertise, and consistency.

This Article concludes by looking briefly at the Facebook Oversight Board, an independent institution that Facebook is currently creating to provide users with transparency and procedure around content moderation. Ideally, such a Board will provide users with explanation and consistency around how the concepts of newsworthiness and public figures are applied. This institution may ultimately result in users gaining procedural and structural protections in the new, private system regulating their expression. The Article concludes that more platforms should adopt such oversight mechanisms to protect users' speech rights and provide accountability, transparency, and due process.

I. PUBLIC FIGURES AND NEWSWORTHINESS IN OLD GOVERNANCE: TORT LAW AND THE FIRST AMENDMENT IN THE COURTS

Within the vast world of tort law, which aims to provide relief when people are harmed, are what Jack Balkin calls "communications torts."[10] Communications torts are "a category of legal causes of action in which people are harmed by speech acts of others."[11] The concept of communications torts is particularly salient in the digital age because, as Balkin foresaw, "all activity in virtual worlds must begin as a form of speech" such that "[w]hen people injure each other in virtual worlds in ways that the law will recognize, they are almost always committing some form of communications tort."[12]

Unsurprisingly, then, legal claims involving communications torts will often implicate the First Amendment because they are premised on harms caused by speech. When adjudicating these claims, the Supreme Court has looked at whether the claimant is a public figure or whether the underlying speech is newsworthy. This Part focuses on how the concepts of public figures and newsworthiness have curtailed three communications torts: defamation, invasion of privacy, and intentional infliction of emotional distress. By unfolding the history and underlying rationales of the Court's First Amendment doctrine in this area, this Part provides context for Part II's exploration of how Facebook developed its own rules surrounding these same concepts. As we will see in Part III, although these public and private

10. Jack M. Balkin, *Law and Liberty in Virtual Worlds*, 49 N.Y.L. SCH. L. REV. 63, 73 (2004).
11. *Id.*
12. *Id.*

doctrines evolved in seemingly disparate contexts, they share many similarities—and, in a new era of online speech, some critical differences.[13]

A. DEFAMATION

The tort of defamation has played a central role in the development of First Amendment law. Although liability for false speech that injures someone's reputation was long thought to raise no constitutional concern, the Supreme Court eventually developed an intricate web of rules to restrain the reach of defamation and protect free speech. The story begins in March 1960 when L.B. Sullivan, an elected commissioner from Alabama, sued the *New York Times* for defamation after the newspaper published an advertisement criticizing the Montgomery Police Department's treatment of civil-rights demonstrators.[14] Sullivan claimed that the advertisement contained falsehoods that damaged his reputation. His case made it all the way to the Supreme Court.[15]

In addressing the threshold issue in *New York Times Co. v. Sullivan*,[16] Justice William Brennan explained that torts like defamation "can claim no talismanic immunity from constitutional limitations" and "must be measured by standards that satisfy the First Amendment."[17] Although he acknowledged that the advertisement contained falsehoods, he explained that "erroneous statement is inevitable in free debate" and "must be protected if the freedoms of expression are to have the 'breathing space' that they 'need . . . to survive.'"[18] Defamation claims, he said, cannot create an environment that "dampens the vigor and limits the variety of public debate."[19] To address this concern, the Court crafted a special constitutional rule: public figures alleging defamation must prove that the offending statements were made with "actual malice"—that is, with knowledge that the statement "was false or with reckless disregard of whether it was false or not."[20]

---

13. *Id.* at 73–76.

14. LEE LEVINE & STEPHEN WERMIEL, THE PROGENY: JUSTICE WILLIAM J. BRENNAN'S FIGHT TO PRESERVE THE LEGACY OF *NEW YORK TIMES V. SULLIVAN* 3–4 (2014).

15. *See generally id.* (providing an excellent account of this historic case); KERMIT L. HALL & MELVIN I. UROFSKY, *NEW YORK TIMES V. SULLIVAN*: CIVIL RIGHTS, LIBEL LAW, AND THE FREE PRESS (2011) (same); Mary-Rose Papandrea, *The Story of* New York Times Co. v. Sullivan, *in* FIRST AMENDMENT STORIES 229 (Richard W. Garnett & Andrew Koppelman eds., 2012) (same).

16. N.Y. Times Co. v. Sullivan, 376 U.S. 254 (1964).

17. *Id.* at 269.

18. *Id.* at 271–72 (first alteration in original) (quoting NAACP v. Button, 371 U.S. 415, 433 (1963)).

19. *Id.* at 279.

20. *Id.* at 279–80.

The justices in *Sullivan* identified two justifications for creating this constitutional hurdle for public figures. The central reason—one that Justice Brennan discussed at length in the majority opinion—was the democratic imperative of preserving "debate on public issues."[21] This rationale reflects a theory of the First Amendment grounded in democratic self-governance. Various scholars have shaped the contours of this theory and glossed it in different ways,[22] but a central concern for many of them is the public's need to have the information necessary to engage in self-government.[23] Under this rationale, the First Amendment protects the public's entitlement "to all information that is necessary for informed governance" because "the public, in its role as the electorate, [is] ultimately responsible for political decisions."[24] The self-governance theory is often cached out in these "educative" terms by justifying "speech protection not because of any individual right of expression but instead because of the need to create an informed public."[25]

---

21. *Id.* at 270.

22. *Compare, e.g.*, ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (1948) (developing a listener-focused account for the relationship between free speech and self-government through the idea of the educative function of a town meeting), *with* ROBERT C. POST, CONSTITUTIONAL DOMAINS: DEMOCRACY, COMMUNITY, MANAGEMENT (1995) (developing a speaker-focused account that connects free speech and self-government by exploring the legitimating function served by public discourse).

23. *See generally* Thomas E. Kadri, *Drawing Trump Naked: Curbing the Right of Publicity to Protect Public Discourse*, 78 MD. L. REV. 899, 905 (2019) (discussing theorists who hold this view and dubbing them "educative" theorists).

24. Robert C. Post, *The Social Foundations of Privacy: Community and Self in the Common Law Tort*, 77 CALIF. L. REV. 957, 999 (1989).

25. Kadri, *supra* note 23, at 913; *see also* OWEN M. FISS, LIBERALISM DIVIDED: FREEDOM OF SPEECH AND THE MANY USES OF STATE POWER 13 (1996) (attesting that "[w]e allow people to speak so others can vote" because "[s]peech allows people to vote intelligently and freely, aware of all the options and in possession of all the relevant information"); ALEXANDER MEIKLEJOHN, POLITICAL FREEDOM: THE CONSTITUTIONAL POWERS OF THE PEOPLE 55 (1960) (arguing that the First Amendment "has no concern about the 'needs of many men to express their opinions'" but rather is concerned with "the common needs of all the members of the body politic"); *id.* at 56–57, 61 (criticizing Zechariah Chafee, Jr.'s "inclusion of an individual interest within the scope of the First Amendment," and Justice Oliver Wendell Holmes's "excessive individualism" on this front); Owen M. Fiss, *Free Speech and Social Structure*, 71 IOWA L. REV. 1405, 1409–10 (1986) (arguing that "[t]he purpose of free speech is not individual self-actualization, but rather the preservation of democracy, and the right of a people, as a people, to decide what kind of life it wishes to live"); Owen M. Fiss, *Why the State?*, 100 HARV. L. REV. 781, 786 (1987) (framing the individual speech right in instrumental terms, worthy of protection "only when it enriches public debate"); Cass R. Sunstein, *Television and the Public Interest*, 88 CALIF. L. REV. 499, 501 (2000) (maintaining that the primary purpose of free speech is to promote deliberative democracy— "a system in which citizens are informed about public issues and able to make judgments on the basis of reasons"). For some skeptical treatment of these theorists, see J.M. Balkin, *Populism and Progressivism as Constitutional Categories*, 104 YALE L.J. 1935, 1935–90 (1995); Robert Post, *Meiklejohn's Mistake: Individual Autonomy and the Reform of Public Discourse*, 64 U. COLO. L. REV. 1109, 1109–23 (1993).

Today, *Sullivan* is often framed as a case that deals solely with the importance of speech about "public figures," but this undervalues its significance to a much broader doctrine. Justice Brennan's concern was not limited to protecting speech about a plaintiff's political position as a government official, but rather what the justice called a "profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open."[26] The focus was not simply on Sullivan's social status, but also—at least implicitly—on whether the speech at issue related to issues that the public needed to know in order to self-govern. The question whether a defamation plaintiff is a public figure is part of a subtler inquiry about whether the speech in question touches on a matter of public concern.[27] As a result, an important part of *Sullivan*'s legacy is that it weaved self-governance theory into First Amendment doctrine.

Although the self-governance theory animated the Court's opinion in *Sullivan*, a second rationale for treating public figures differently appeared in Justice Arthur Goldberg's concurrence: public figures enjoy "equal if not greater access than most private citizens to media of communication."[28] Though not part of the majority's reasoning, this additional justification eventually animated the development of public-figure doctrine. When a university athletic director and a prominent political activist sued two newspapers for defamation in *Curtis Publishing Co. v. Butts*, the Court adopted Justice Goldberg's rationale by justifying its decision on the basis that plaintiffs "had sufficient access to the means of counterargument" to rebut the alleged falsehoods.[29] The Court also echoed the self-governance rationale in explaining why *Sullivan*'s constitutional rule applied to plaintiffs who were not government officials,[30] explaining that the "public interest" in being informed about nonpolitical public figures was "not less" than being informed about public officials.[31] Thus, although the Court's holding hinged on the plaintiffs' social status, the justification for it was again rooted in the importance of robust debate on "public issues,"[32] as well as the fact that the plaintiffs could participate in that debate by virtue of their prominence.

---

26. *Sullivan*, 376 U.S. at 270

27. *See* Catherine Hancock, *Origins of the Public Figure Doctrine in First Amendment Defamation Law*, 50 N.Y.L. SCH. L. REV. 81, 85 (2005); Post, *supra* note 24, at 997.

28. *Sullivan*, 376 U.S. at 304–05 (Goldberg, J., concurring).

29. Curtis Publ'g Co. v. Butts, 388 U.S. 130, 135–41, 155 (1967).

30. *Id.* at 155.

31. *Id.* at 154–55; *see also id.* at 163 (Warren, C.J., concurring in result) (declaring that distinguishing between public figures and officials would have "no basis in law, logic, or First Amendment policy").

32. *Id.* at 147.

Given the Court's reliance on self-governance theory to protect speech on "public issues," one might wonder why the Court focused on the plaintiff's status instead of simply asking whether the speech at issue was necessary for the public to know. The Court flirted with such a reformulation of the doctrine in *Rosenbloom v. Metromedia, Inc.*, in which a plurality extended *Sullivan*'s rule to all defamation claims involving speech on matters of public concern, regardless of whether the plaintiff was a public or private figure.[33] Justice Brennan's plurality opinion reasoned that a matter "of public or general interest . . . cannot suddenly become less so merely because a private individual is involved, or because in some sense the individual did not 'voluntarily' choose to become involved."[34] In his view, "[t]he public's primary interest is in *the event*," not the individual's social status.[35] Any interest in the individual's prominence was merely a corollary to that primary interest, for "the public focus is on the conduct of the participant and the content, effect, and significance of the conduct, not the participant's prior anonymity or notoriety."[36] In order to "honor the commitment to robust debate on public issues . . . embodied in the First Amendment" that the Court had recognized in *Sullivan*, Justice Brennan concluded that the constitutional rule must apply "to all discussion and communication involving matters of public or general concern, without regard to whether the persons involved are famous or anonymous," though quite what constituted a matter of "public or general concern" was left rather opaque.[37]

Rosenbloom's doctrinal simplicity—if jettisoning the "public figure" concept for the ambiguous notion of the "public or general concern" can be considered simplicity—was fleeting. Just three years later, the Court held in *Gertz v. Robert Welch, Inc.* that *Sullivan*'s rule should not apply to claims brought by private figures.[38] In drawing lines between public and private figures, the Court imagined two, and possibly three, types of public figures: general public figures; limited-purpose public figures; and, perhaps, involuntary public figures.[39] General public figures "occupy positions of such persuasive power and influence that they are deemed public figures for

33. Rosenbloom v. Metromedia, Inc., 403 U.S 29, 43–44 (1971) (plurality opinion).
34. *Id.* at 43.
35. *Id.* (emphasis added).
36. *Id.*
37. *Id.* at 43–44.
38. Gertz v. Robert Welch, Inc., 418 U.S. 323, 339–48 (1974).
39. *See* W. Wat Hopkins, *The Involuntary Public Figure: Not So Dead After All*, 21 CARDOZO ARTS & ENT. L.J. 1, 21 (2003) ("[T]here is disagreement as to whether the Supreme Court identified two or three categories of public figure status.").

all purposes,"[40] while limited-purpose public figures "thrust themselves to the forefront of particular public controversies in order to influence the resolution of the issues involved."[41] Notably, the Court left undefined the third (possible) type of public figure, observing only that "truly involuntary public figures must be exceedingly rare."[42] This remark was prophetic: the Court has never encountered its mythical character in the forty years that have since passed.[43]

The demarcation of different types of public figures has developed into a central concern in the Court's defamation jurisprudence. In contrast to the atrophied idea of the involuntary public figure, the Court has developed the "voluntariness" element at length. In *Gertz*, the Court explained that both general and limited-purpose public figures "invite attention and comment" and thus have "voluntarily exposed themselves to increased risk of injury from defamatory falsehood concerning them."[44] Building from this premise, the Court in *Time, Inc. v. Firestone* held that a woman who divorced a member of a wealthy socialite family was not a public figure.[45] The Court cautioned against adopting a concept of public figures that might create too large a class, and explained that a person needed more than local notoriety to become a public figure.[46] The Court stressed that the divorcée did not "freely choose to publicize issues as to the propriety of her married life" but rather was "compelled" to go to court to end her marriage.[47] Her actions, then, were "no more voluntary in a realistic sense" than those of a criminal defendant "called upon to defend his interests in court."[48] In addition, the Court refused to equate "public controversy" with "all controversies of interest to the public" because doing so would reinstate the *Rosenbloom* rule that disavowed the significance of a plaintiff's social status.[49]

Since *Firestone*, the Court has repeatedly held that individuals who did not *voluntarily* garner attention were not public figures. In *Wolston v. Reader's Digest Association*, for example, the Court declined to apply the

---

40. *Gertz*, 418 U.S. at 345.
41. *Id.*
42. *Id.*
43. *See* Jeffrey Omar Usman, *Finding the Lost Involuntary Public Figure*, 2014 UTAH L. REV. 951, 952 (claiming that "involuntary public figures" as category of individuals in First Amendment defamation jurisprudence "has become lost").
44. *Gertz*, 418 U.S. at 345.
45. Time, Inc. v. Firestone, 424.U.S. 448, 453 (1976).
46. *Id.* at 450–53.
47. *Id.* at 454.
48. *Id.* (citation omitted).
49. *Id.*

label to a witness who missed a grand-jury hearing investigating Cold War espionage.[50] The witness had not "voluntarily thrust" or "injected" himself into the public eye; rather, the Court declared that "[i]t would be more accurate to say that [he] was dragged unwillingly into the controversy."[51] Similarly, in *Hutchinson v. Proxmire*, in which a professor sued a U.S. Senator who criticized him for wasting federal funds, the Court stressed that the professor remained a private figure because he had not "thrust himself or his views into public controversy to influence others."[52] The Court rejected the lower courts' conclusion that the professor became a public figure in part because he gained some prominence and access to the media *after* the Senator had allegedly defamed him.[53] Although various media outlets reported the professor's response to the Senator's criticism, the Court stressed that the professor "did not have the regular and continuing access to the media that is one of the accouterments of having become a public figure."[54] In other words, the controversy itself could not transform the professor into a public figure simply because he appeared in the news as a result.

As these cases show, although the Court glossed the public-figure doctrine with a new taxonomy in *Gertz* and its progeny, the basic rationales for affording heightened constitutional protection for speech about public figures tracked the pair of rationales originally suggested in *Sullivan* and *Butts*. First, speech about public figures requires protection for "debate on public issues [to] be uninhibited, robust, and wide-open."[55] The pervasive power of some people makes their behaviors a matter of public interest, just as the behaviors of someone who thrusts herself into a public controversy become a matter of public interest, and the public needs to know about these behaviors if it is to engage meaningfully in self-governance. Second, public figures are "less vulnerable to injury from defamatory statements" because they generally have greater ability to engage in "self-help" by "counter[ing] criticism and expos[ing] the falsehood and fallacies of defamatory statements" in the media.[56] These dual rationales sustain much of modern public-figure doctrine to this day.[57]

---

50. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 166–67 (1979).
51. *Id.* at 166.
52. Hutchinson v. Proxmire, 443 U.S. 111, 135 (1979).
53. *Id.* at 134–36.
54. *Id.* at 136.
55. N.Y Times Co. v. Sullivan, 376 U.S. 254, 270 (1964).
56. *Wolston*, 443 U.S. at 164.
57. *See generally* Shlomit Yanisky-Ravid & Ben Zion Lahav, *Public Interest vs. Private Lives—*

It's crucial to note, however, that the Court's doctrine also reflects two caveats to these rationales. The first caveat relates to the significance of the idea of "voluntariness." The idea that a person *chooses* to place herself in the public eye underlies the "normative consideration"[58] that public figures are "less deserving"[59] of protection from defamation because, unlike private figures, they have "voluntarily exposed themselves to increased risk of injury from defamatory falsehood."[60] In striking the balance between protecting free speech and remedying harmful speech, the Court has continued to cling to the importance of voluntariness, even when the idea of voluntariness potentially conflicts with the public-figure doctrine's embrace of self-governance theory and its assertion that private figures lack the means of rebuttal. After all, the fact that a plaintiff freely chose to enter the public arena might not mean that he or she commands more public interest or that he or she has greater access to the media—private figures might get caught up in events that raise issues of great importance to the public, and the attention that flows from these events might furnish these private figures with ample opportunity to respond in the media. Nonetheless, the Court has embraced this doctrinal tension to date.

The second caveat to the rationales developed in *Sullivan* and *Butts* concerns the *Rosenbloom* retraction in *Gertz*. The *Rosenbloom* plurality had sought to abandon distinctions based on social status,[61] and even based on voluntariness,[62] but the Court in *Gertz* clawed back the significance of both distinctions in the constitutional analysis. Had the Court's sole concern been to preserve "debate on public issues," *Rosenbloom*'s rule would have carried the day, for Justice Brennan was surely right when he observed in *Rosenbloom* that matters of public concern "cannot suddenly become less so merely because a private individual is involved, or because in some sense the individual did not 'voluntarily' choose to become involved."[63] Yet despite this fact, the Court chose not to embrace a more speech-protective rule that would cover all speech on matters of public concern. *Gertz* and later cases recognized the fundamental importance of free speech, but the Court

---

Affording Public Figures Privacy in the Digital Era: The Three Principle Filtering Model, 19 J. CONST. L. 975, 983–84 (2017) (outlining the various components of the Court's public-figure doctrine as (1) "access and control over the media"; (2) "enrollment in a special role in the public eye"; (3) "willingly (voluntarily) choosing to engage in a public role, inviting invasion of privacy risks"; and (4) "public controversy").

58. *Wolston*, 443 U.S. at 164.

59. *Id.*

60. *Id.* (quoting Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974)).

61. Rosenbloom v. Metromedia, Inc., 403 U.S 29, 43–44 (1971).

62. *Id.* at 43.

63. *Id.* at 43–44.

has nonetheless insisted upon a different constitutional regime for private figures.

These two caveats reveal something important about the balance that the Court has struck between preserving free speech and protecting people from harmful speech. In the context of false and defamatory speech, the Court has developed a constitutional commitment that the First Amendment must limit tort liability in order to protect robust public debate. But the Court has also recognized that the public's eagerness to engage in such debate is not always sufficient to override all attempts to tackle harmful speech through defamation law. As the next section reveals, the Court has struck a different balance when faced with other tort claims that raise free-speech concerns.

B. INVASION OF PRIVACY

Given the distinctions drawn in defamation law between *private* and *public* figures and matters of *private* and *public* concern, it is unsurprising that privacy law has attracted similar constitutional concepts. Like the defamation tort, privacy torts often involve allegations that a speech act has caused harm to someone, and thus these claims regularly implicate the First Amendment. After *Sullivan* constitutionalized defamation law through the public-figure doctrine, defendants in privacy actions began raising arguments that similar limitations should be placed on privacy torts in order to preserve free speech. The resulting doctrine borrows heavily from the Court's defamation jurisprudence but differs in important respects.

Just a few years after *Sullivan*, the Court applied "the First Amendment principles pronounced in [*Sullivan*]" within the privacy realm in *Time, Inc. v. Hill*,[64] partially migrating *Sullivan*'s rule outside the defamation context for the first time. The case developed after James Hill and his family were held hostage by escaped convicts.[65] Hill sued *Life Magazine* after it published an article suggesting that a new play portrayed his family's story.[66] Although he maintained that the article was "false and untrue,"[67] his claim sounded not in defamation but in the privacy tort of unreasonably placing a person in a "false light" before the public.[68] Hill argued that he was not a public figure but a private citizen who had "involuntarily become

64. Time, Inc. v. Hill, 385 U.S. 374, 390 (1967); *see also* Hancock, *supra* note 27, at 105–12 (discussing the relationship of *Hill* to *Sullivan*).

65. *Hill*, 385 U.S. at 378.

66. *Id.*

67. *Id.*

68. *See* RESTATEMENT (SECOND) OF TORTS §§ 652A, 652E (AM. LAW INST. 1977).

newsworthy" after he and his family were victims of a crime.[69]

Justice Brennan again delivered the Court's opinion, stressing—as he had done in *Sullivan*—that "[f]reedom of discussion . . . must embrace all issues about which information is needed."[70] Foreshadowing an issue that would become crucial in the Court's defamation jurisprudence, Justice Brennan declined to base a constitutional rule on "the distinction . . . between the relative opportunities of the public official and the private individual to rebut" harmful speech.[71] Indeed, he eschewed consideration of the plaintiff's social status entirely, as he would later attempt to do in *Rosenbloom*. He focused instead on the need to preserve debate on "matters of public interest"[72] and, in so doing, "declared an expansive view of the First Amendment as protection for all newsworthy material,"[73] regardless of whether the plaintiff is a private figure.[74] The decision reflected the self-governance theory of the First Amendment, justified as it was by the fact that the Hill's ordeal concerned "issues about which information is needed" for the public to govern itself effectively.[75]

Hill gave constitutional weight to an idea that had deep foundations in privacy law. Samuel Warren and Louis Brandeis's seminal work advocating for a vigorous right to privacy nonetheless stressed that such a right should not prohibit speech on matters "of public or general interest,"[76] and the first decision acknowledging a right of privacy contained a similar qualification.[77] Sometimes referred to as speech that's "newsworthy,"[78]

---

69. *See* Harry Kalven, Jr., *The Reasonable Man and the First Amendment:* Hill, Butts *and* Walker, 1967 SUP. CT. REV. 267, 279.

70. *Hill*, 385 U.S. at 388 (citation omitted).

71. *Id.* at 391.

72. *See id.* at 387–88 ("We hold that the constitutional protections for speech and press preclude the application of the New York statute to redress false reports of matters of public interest in the absence of proof that the defendant published the report with knowledge of its falsity or in reckless disregard of the truth.").

73. Samantha Barbas, *When Privacy Almost Won:* Time, Inc. v. Hill, 18 U. PA. J. CONST. L. 505, 508 (2015).

74. *Hill*, 385 U.S. at 387–88.

75. *Id.* at 388 (citation omitted).

76. Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 214 (1890).

77. Pavesich v. New England Life Ins., 50 S.E. 68, 74 (Ga. 1905) ("The truth may be spoken, written, or printed about all matters of a public nature, as well as matters of a private nature in which the public has a legitimate interest."); *see also* Post, *supra* note 24, at 996.

78. Time, Inc. v. Hill, 385 U.S. 374, 400 (1967) (Black, J., concurring); *see also* FLA. STAT. § 90.5015(1)(b) (2017) ("'News' means information of public concern relating to local, statewide, national, or worldwide issues or events."); Kadri, *supra* note 23, at 912–13; Mary-Rose Papandrea, *Citizen Journalism and the Reporter's Privilege*, 91 MINN. L. REV. 515, 578–81 (2007) (observing that the newsworthiness standard "involves essentially the same inquiry as a 'public concern' test").

speech on matters of public concern is said to lie "at the core of the First Amendment"[79] for reasons that self-governance theorists have detailed at length.[80]

The privacy tort of public disclosure of private facts has also been cabined by constitutional concerns for free speech. As with the false-light privacy claim in *Hill*, there can be no liability under the disclosure tort when the speech in question is of "legitimate public concern."[81] But while the *Hill* Court seemed ambivalent to the plaintiff's status, courts have superficially entertained a distinction between private and public figures who bring disclosure claims. As an initial matter, this is because people live much of their lives out in the open; as a result, the disclosure tort will often be inapplicable because many facts about people are not truly "private."[82] A complication arises, however, when people seek to shield certain aspects of their lives from public view: whereas private figures might rely on the disclosure tort, the "legitimate interest of the public" may extend "to some reasonable extent" to facts about public figures "that would otherwise be purely private."[83]

Although courts recognize the concept of the public figure under the disclosure tort, they do not discriminate between the voluntary and involuntary public figure as they do in defamation law. The Restatement does draw a distinction between voluntary and involuntary public figures, but it is a distinction without a doctrinal difference.[84] The involuntary public

79. Frisby v. Schultz, 487 U.S. 474, 479 (1988); *see also* NAACP v. Claiborne Hardware Co., 458 U.S. 886, 913 (1982) ("[E]xpression on public issues 'has always rested on the highest rung of the hierarchy of First Amendment values.'" (citation omitted)); Schenck v. Pro-Choice Network of W. N.Y., 519 U.S. 357, 377 (1997) ("Leafletting and commenting on matters of public concern are classic forms of speech that lie at the heart of the First Amendment . . . .").

80. *See, e.g.*, CASS R. SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH 121–65 (1993); Robert H. Bork, *Neutral Principles and Some First Amendment Problems*, 47 IND. L.J. 1, 20–35 (1971); Alexander Meiklejohn, *The First Amendment Is an Absolute*, 1961 SUP. CT. REV. 245, 255.

81. RESTATEMENT (SECOND) OF TORTS § 652D cmt. d (AM. LAW INST. 1977) (noting that this rule applies as a matter of both common law and constitutional law).

82. *Id.* § 652D cmt. e.

83. *Id.* § 652D cmt. f; *see also id.* § 652D cmt. h (noting that interest about a public figure "may legitimately extend, to some reasonable degree, to further information concerning the individual and to facts about him, which are not public and which, in the case of one who had not become a public figure, would be regarded as an invasion of his purely private life").

84. A voluntary public figure "voluntarily places himself in the public eye, by engaging in public activities, or by assuming a prominent role in institutions or activities having general economic, cultural, social or similar public interest, or by submitting himself or his work for public judgment." *Id.* § 652D cmt. e. Involuntary public figures, by contrast, "have not sought publicity or consented to it, but through their own conduct or otherwise have become a legitimate subject of public interest"—"[t]hey have, in other words, become 'news.'" *Id.* § 652D cmt. f.

figure is an enigmatic and elusive character in defamation law,[85] but in privacy law she is pervasive: she is the person who commits a crime or is accused of it,[86] she is the victim or witness of crime and catastrophe,[87] and she is even the hapless soul who inadvertently gets caught up in "events that attract public attention."[88] Thus, even if she does nothing to "thrust" herself voluntarily into the vortex of a public controversy, the involuntary public figure is "properly subject to the public interest" and subject to the same constitutional rules as those who freely enter the public arena.[89] Her desire to avoid the limelight is irrelevant, as is her lack of access to the media as a means of engaging in counter-speech.[90]

This complex web of considerations creates a paradoxically simple rule: there is no privacy invasion when the fact disclosed is of "legitimate public concern." This is true for disclosures about private figures and disclosures about public figures, whether they be voluntary or involuntary. The only wrinkle is that some facts about public figures would not be of "legitimate public concern" were they instead about private figures. But all that means is that certain facts about public figures—even facts that would otherwise be "private"—are matters of legitimate public concern *because of the person's social status*. The court's status determination, then, serves simply as a proxy for deciding whether a particular fact is of legitimate public concern—that is, whether it implicates the public's ability to engage in self-governance.

This means that defining "legitimate public concern" carries a lot of analytical weight; and indeed, a privacy claim will often rise and fall on this determination alone. One might think that the definition is both descriptive and circular: a matter is of legitimate public concern when the public is concerned about the matter. The Restatement gestures at this conception when it observes that "[i]ncluded within the scope of legitimate public concern are matters of the kind customarily regarded as 'news'" or matters that have "popular appeal."[91] This descriptive definition means that, "[t]o a considerable extent, . . . publishers and broadcasters have themselves

85. *See supra* text accompanying notes 21–25.

86. *See, e.g.*, Miller v. Nat'l Broad. Co., 157 F. Supp. 240, 241 (D. Del. 1957); RESTATEMENT (SECOND) OF TORTS § 652D cmt. f. (noting that criminals are involuntary public figures because they "may not only not seek publicity but may make every possible effort to avoid it").

87. *See, e.g.*, Jones v. Herald Post Co., 18 S.W.2d 972, 972–73 (Ky. 1929); RESTATEMENT (SECOND) OF TORTS § 652D cmt. f.

88. RESTATEMENT (SECOND) OF TORTS § 652D cmt. f; *see, e.g.*, Jacova v. S. Radio & Television Co., 83 So. 2d 34, 40 (Fla. 1955).

89. RESTATEMENT (SECOND) OF TORTS § 652D cmt. f.

90. *See id.*

91. *Id.* § 652D cmt. g.

defined the term, as a glance at any morning paper will confirm."[92] In short, the reporting of a fact in the news is highly probative, if not conclusive, of its newsworthiness.

Broad deference to the news media is an idea with some constitutional pedigree. Given that many privacy claims are brought against the press, the vaulted status of speech on matters of public concern has sometimes been called the "privilege to report news" or the "privilege to publicize newsworthy matters."[93] In *Cox Broadcasting Corp. v. Cohn,* the Court barred a disclosure claim brought against a media company that reported a rape victim's name obtained from public court records.[94] In holding that "the press cannot be sanctioned for publishing" information found in public documents, the Court stressed that "reliance must rest upon the judgment of those who decide what to publish or broadcast."[95]

Despite the sweeping language in some judicial opinions, the reality is that many courts have shunned a purely descriptive definition of newsworthiness that would yield entirely to the press. The deference is broad but not absolute.[96] Much hinges on whether the public's interest in knowing a particular fact is "legitimate." The legitimacy determination, in turn, considers "the customs and conventions of the community," meaning that "what is proper becomes a matter of the community mores."[97] In the language of the Restatement:

> The line is to be drawn when the publicity ceases to be the giving of information to which the public is entitled, and becomes a morbid and sensational prying into private lives for its own sake, with which a reasonable member of the public, with decent standards, would say that he had no concern. The limitations, in other words, are those of common decency, having due regard to the freedom of the press and its reasonable leeway to choose what it will tell the public, but also due regard to the

92. *Id.*

93. Post, *supra* note 24, at 995 (quoting, respectively, Harry Kalven, Jr., *Privacy in Tort Law—Were Warren and Brandeis Wrong?*, 31 LAW & CONTEMP. PROBS. 326, 336 (1966) and Virgil v. Time, Inc., 527 F.2d 1122, 1128 (9th Cir. 1975)).

94. Cox Broad. Corp. v. Cohn, 420 U.S. 469, 470, 496–97 (1975).

95. *Id.* at 496; *see also* Erin C. Carroll, *Making News: Balancing Newsworthiness and Privacy in the Age of Algorithms*, 106 GEO. L.J. 69, 77–81 (2017).

96. *See, e.g.*, RESTATEMENT (SECOND) OF TORTS § 652D cmt. h ("The extent of the authority to make public private facts [about public figures] is not, however, unlimited.").

97. *Id.*; *see also id.* § 652D cmt. g (explaining that the media's broad leeway to define newsworthiness must still accord "with the mores of the community"); Amy Gajda, *The Present of Newsworthiness*, 50 NEW ENG. L. REV. 145, 145–46 (2016).

> feelings of the individual and the harm that will be done to him by the exposure.[98]

These principles were at the heart of a recent blockbuster privacy lawsuit between former wrestler Hulk Hogan and the now-defunct news organization Gawker. Hogan sued for invasion of privacy after Gawker published an excerpted sex tape that showed him ensconced with his best friend's wife. Although multiple courts refused to enjoin publication of the tape on the grounds that it was newsworthy,[99] a jury disagreed and awarded Hogan massive damages.[100] Evidently, the jury concluded that publishing the tape amounted to "morbid and sensational prying" into Hogan's private life that violated "common decency," despite Hogan's clear status as a public figure.[101]

C. INTENTIONAL INFLICTION OF EMOTIONAL DISTRESS

The constitutional concepts of public figures and newsworthiness have developed in relation to a third tort: intentional infliction of emotional distress ("IIED"). Under this tort, a plaintiff must show that the defendant "intentionally or recklessly engaged in extreme and outrageous conduct that caused the plaintiff to suffer severe emotional distress."[102] When the "extreme and outrageous conduct" consists of actions protected by the First Amendment, courts have crafted rules to limit the tort's incursion on free speech.

The Court's first tussle with the IIED tort came in *Hustler Magazine, Inc. v. Falwell*, a legal battle fit—and indeed destined[103]—for Hollywood.[104] Jerry Falwell, a nationally renowned Christian minister, sued *Hustler Magazine* and its antagonistic publisher, Larry Flynt, after the magazine spoofed an interview with Falwell in a liquor advertisement entitled "Jerry Falwell talks about his first time."[105] A jury accepted Falwell's claim that the

---

98. RESTATEMENT (SECOND) OF TORTS § 652D cmt. h.

99. For background on the case, see generally Bollea v. Gawker Media, LLC, 913 F. Supp. 2d 1325 (M.D. Fla. 2012); Gawker Media, LLC v. Bollea, 170 So. 3d 125 (Fla. Dist. Ct. App. 2015); Gawker Media, LLC v. Bollea, 129 So. 3d 1196 (Fla. Dist. Ct. App. 2014).

100. Nick Madigan & Ravi Somaiya, *Hulk Hogan Awarded $115 Million in Privacy Suit Against Gawker*, N.Y. TIMES (Mar. 18, 2016), https://www.nytimes.com/2016/03/19/business/media/gawker-hulk-hogan-verdict.html [https://perma.cc/EEZ3-NYMF].

101. *See* RESTATEMENT (SECOND) OF TORTS § 652D cmt. h; *see also* Garner v. Triangle Publ'ns, 97 F. Supp. 546, 549–50 (S.D.N.Y. 1951).

102. *See* Snyder v. Phelps, 562 U.S. 443, 451 (2011).

103. THE PEOPLE VS. LARRY FLYNT (Phoenix Pictures 1996).

104. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 50 (1988).

105. *Id.* at 48.

parody caused him grave emotional harm, awarding him compensatory and punitive damages, but the Court resoundingly rejected Falwell's claim and extended *Sullivan*'s rule to IIED claims brought by public figures.[106] Justice William Rehnquist wrote for a unanimous Court that public figures must satisfy the rigors of actual malice. The decision was grounded in the *Sullivan* and *Hill* self-governance rationales about the need for "robust political debate" and "the free flow of ideas and opinions on matters of public interest and concern."[107]

Portions of the *Hustler* decision stressed the significance of Falwell's social status as a public figure. The Court spoke, for instance, of the First Amendment right to be "critical of those who hold public office or those public figures who are 'intimately involved in the resolution of important public questions or, by reason of their fame, shape events in areas of concern to society at large.'"[108] But this focus on social status faded away the next time the Court addressed the constitutionality of an IIED claim. In *Snyder v. Phelps*, the father of a soldier killed in Iraq sued parishioners from the Westboro Baptist Church who protested near his son's funeral, with an array of signs, including "God Hates Fags," "Thank God for Dead Soldiers," and "Priests Rape Boys."[109] The father's leading arguments in the Supreme Court revolved around the fact that he was a private figure who "took no action to inject himself into a public debate" and "did nothing to obtain the status of a celebrity or a public figure."[110]

The Court, however, was unpersuaded. Rather than focusing on the father's social status, the Court held that the First Amendment's application to the father's claim turned on whether the parishioners' speech was "of public or private concern."[111] Chief Justice John Roberts explained that the First Amendment protections are "less rigorous" when speech regulations target matters of *private* concern because "[t]here is no threat to the free and robust debate of public issues; there is no potential interference with a meaningful dialogue of ideas; and the threat of liability does not pose the risk

106. *Id.* at 48, 56.

107. *Id.* at 50–51; *see also supra* notes 16–27 & 64–74 and accompanying text.

108. *Hustler*, 485 U.S. at 51 (quoting Curtis Publ'g Co. v Butts, 388 U.S. 130, 164 (1967) (Warren, C.J., concurring in result)).

109. Snyder v. Phelps, 562 U.S. 443, 448 (2011).

110. Brief for Petitioner at 34, Snyder v. Phelps, 562 U.S. 443 (2011) (No. 09-751); *see also* Reply Brief at 10–14, Snyder v. Phelps, 562 U.S. 443 (2011) (No. 09-751) (arguing at length that Snyder was a private figure, not a public one).

111. *Snyder*, 562 U.S at 451.

of a reaction of self-censorship on matters of public import."[112] In short, the Court cared only about whether the underlying speech concerned issues that the public needed to know in order to govern itself. Given the analytical importance of determining whether speech is on a matter of public concern, one might think that the Court would clearly define this constitutional concept. Yet the Court candidly admitted that its boundaries "are not well defined"[113] before describing it in disjunctive—and potentially contradictory—terms as speech that is (1) "fairly considered as relating to any matter of political, social, or other concern to the community," or (2) "a subject of legitimate news interest; that is, a subject of general interest and of value and concern to the public."[114] The definition might be descriptive (what the public *does* know or *wants* to know) or normative (what the public *ought* to know or *needs* to know). As applied to the father's claim, the Court concluded that the parishioners' signs highlighted issues that are "matters of public import" and as such gained First Amendment protection that blocked liability under the IIED tort.[115] The Court briefly raised the issue of directionality, suggesting that at least two of the signs—"You're Going to Hell" and "God Hates You"—could be "viewed as containing messages related to Matthew Snyder or the Snyders specifically."[116] The fact that the Court even flagged the issue of directionality might imply that speech targeting particular people raises different constitutional considerations, but the Court never reached this question because "the overall thrust and dominant theme" of the signs "spoke to broader public issues."[117]

Snyder, the Court's most recent case concerning communication torts and the First Amendment, reveals several important points about the constitutional dynamics at play when the Court sets rules for defamation, privacy, and IIED claims. All three regimes reflect the Court's concern for protecting robust public discourse—a concern animated by a self-governance theory of the First Amendment. This common thread runs through the jurisprudence for all three torts, even as the rules differ slightly between them. The Court does not disparage the state's interests in protecting people from harmful speech, nor does it dispute that the speech at issue in these cases in fact inflicted harm. Rather, the Court frames the various constitutional rules in prophylactic terms—as creating the conditions

112. *Id.* at 452 (internal quotation marks omitted) (quoting Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., 472 U.S. 749, 760 (1985)).

113. *Id.* at 452 (quoting City of San Diego v. Roe, 543 U.S. 77, 83 (2004) (per curiam)).

114. *Id.* at 453 (citations omitted).

115. *Id.* at 454.

116. *Id.*

117. *Id.*

necessary for an ecosystem in which free speech can flourish. As we'll see, Facebook has viewed its own rulemaking in strikingly similar terms.

II. PUBLIC FIGURES AND NEWSWORTHINESS IN NEW GOVERNANCE: CONTENT MODERATION AND FREE SPEECH AT FACEBOOK

Facebook is the preeminent social network of the digital age. With over two billion users, the platform hosts vast amounts of content shared by people scattered across the globe.[118] Though we might conceive of Facebook as a "New Governor" because of its power over online discourse, it is, of course, a private company that need not satisfy the First Amendment when policing its users' speech.[119] Nor is Facebook bound by any "Constitution" of its own.[120] Instead, Facebook implements a system of semi-public rules called "Community Standards,"[121] which are effectively Facebook's "laws" that govern what users may say on the platform.[122]

In order to implement the Community Standards, Facebook has developed an immense bureaucratic system to moderate user content and adjudicate disputes arising from that content. Because an enormous volume of content is posted every day, Facebook cannot and does not proactively police all violations of its rules. Automated detection of violations is quite sophisticated and successful for various types of visual content (such as child pornography) but less so for written content that poses "nuanced linguistic challenges" (such as harassment and hate speech).[123] As a result, the platform still relies on users to reactively flag speech that might violate its rules.

---

118. *Company Info*, FACEBOOK NEWSROOM, https://newsroom.fb.com/company-info [https://perma.cc/G7SJ-V4E2].

119. *See* Klonick, *supra* note 5, at 1658–62. The question whether the First Amendment restricts the *government's* use of social media presents discrete doctrinal challenges that others have expertly analyzed. *See* Lyrissa Lidsky, *Public Forum 2.0*, 91 B.U. L. REV. 1975, 1979–2002 (2011) (discussing how the public-forum doctrine might apply to when government actors use social media); Helen Norton & Danielle Keats Citron, *Government Speech 2.0*, 87 DENV. U. L. REV. 899, 899 (2010) (analyzing how the government-speech doctrine might adapt given that government's increasing reliance on social media).

120. Though perhaps it should be. *See* Klonick & Kadri, *supra* note 5.

121. *See Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards [https://perma.cc/J7BQ-96E2].

122. DAVID KAYE, SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET 23 (2019).

123. *See* Mark Zuckerberg, *A Blueprint for Content Governance and Enforcement*, FACEBOOK (Nov. 15, 2018), https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634 [https://perma.cc/P4H6-6HVQ]; *see also Community Standards Enforcement Report*, FACEBOOK (Nov. 2019), https://transparency.facebook.com/community-standards-enforcement [https://perma.cc/S8FW-TXXV] (breaking down statistics for removal of nine different types of content, including hate speech and child nudity and sexual exploitation).

Content reported by users is placed into an online queue for review by human content moderators—people working either directly for Facebook or as contractors who are trained to apply Facebook's rules and determine if content violates the Community Standards.[124] The platform removes speech found to be in violation; the rest remains.[125]

Facebook's first internal guidelines for content moderation were created largely by Dave Willner in 2009, who then joined Jud Hoffman to spearhead a small team that formalized and consolidated the ad hoc rules that Facebook's earliest content moderators had been using.[126] Ever since, Facebook has devoted considerable attention to the rules and procedures it uses to govern speech on the platform.[127] Somewhat like a common-law legal system,[128] Facebook regularly adapts its Community Standards to address changing circumstances, including new factual scenarios or technologies; criticism or feedback from outside observers; changing norms surrounding particular issues; and interventions from upper-level management.[129] Both Willner and Hoffman were heavily involved in developing Facebook's rules surrounding public figures and newsworthiness—a history we turn to now.[130]

A. CYBERBULLYING

Whereas the Supreme Court's public-figure doctrine emerged from claims of defamation,[131] Facebook's rules surrounding public figures first

---

124. *See* Klonick, *supra* note 5, at 1630–48. For pathbreaking work on content moderation, particularly on the experiences of the content moderators themselves, see generally SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA (2019). And for excellent exploration of platform policies from sociological and legal perspectives, see generally TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA (2018), and NICHOLAS SUZOR, LAWLESS: THE SECRET RULES THAT GOVERN OUR DIGITAL LIVES (2019).

125. *See* Klonick, *supra* note 5, at 1630–48.

126. Telephone Interview with Jud Hoffman, Former Glob. Policy Manager, Facebook (Jan. 22, 2016); Telephone Interview with Dave Willner, Former Head of Content Policy, Facebook, & Charlotte Willner, Former Safety Manager, User Operations, Facebook (Mar. 23, 2016). All interview notes are on file with the authors.

127. *See* Klonick, *supra* note 5, at 1630–48.

128. *See id.* at 1645–47.

129. *See, e.g.*, Simon Adler, *Post No Evil*, WNYC STUDIOS: RADIOLAB (Aug. 17, 2018), https://www.wnycstudios.org/story/post-no-evil [https://perma.cc/J54X-5CWA]; Sheera Frenkel et al., *Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis*, N.Y. TIMES (Nov. 14, 2018), https://nyti.ms/2DlsGPi [https://perma.cc/MD7T-LVDW].

130. The following Sections contain excerpts from research and commentary discussed in Klonick, *supra* note 5, at 1604–09.

131. *See supra* Section I.A.

developed in response to claims about bullying.[132] In 2009, anti-cyberbullying groups were pressuring Facebook to do more to protect children from online abuse.[133] The problem, however, was that traditional academic definitions of bullying seemed impossible to translate to online content moderation. "How do we write a rule about bullying?" recounted Willner.[134]

> What is bullying? What do you mean by that? It's not just things that are upsetting; it's defined as a pattern of abusive or harassing unwanted behavior over time that is occurring between a higher power [and] a lower power. But that's not an answer to the problem that resides in the content—you can't determine a power differential from looking at the content. You often cannot even do it from looking at their profiles.[135]

The apparent impossibility of employing a traditional definition of bullying meant that Facebook had to make a choice. It could err on the side of keeping up potentially harmful content, or it could err on the side of removing all potential acts of bullying, even if some of the removed content turned out to be benign. Faced with intense pressure from advocacy groups and media coverage on cyberbullying, Facebook opted for the latter approach, but with a caveat. The new presumption in favor of removing speech reported to be "bullying" would apply only to speech directed at private figures. "What we said was, 'Look, if you tell us this is about you, and you don't like it, and you're a private individual—you're not a public figure—then we'll take it down,'" said Hoffman.[136] "Because we can't know whether all of those other elements [of bullying] are met, we had to just make the call to create a default rule for removal of bullying."[137]

Although Hoffman denies borrowing directly from the First Amendment doctrine, his justification for creating this rule tracks some of the reasoning in *Sullivan* and subsequent cases.[138] In order to preserve robust public discourse on the platform, Hoffman's team made the conscious decision to treat certain targets of allegedly harmful speech differently on

---

132. Telephone Interview with Jud Hoffman, Former Glob. Policy Manager, Facebook (Mar. 6, 2018); Telephone Interview with Dave Willner, Former Head of Content Policy, Facebook (Mar. 7, 2018). On the distinct complications of *criminalizing* cyberbullying, see Lyrissa Lidsky & Andrea Pinzon Garcia, *How Not to Criminalize Cyberbullying*, 77 MO. L. REV. 693 (2012).

133. Telephone Interview with Dave Willner, *supra* note 132.

134. *Id.*

135. *Id.*

136. Telephone Interview with Jud Hoffman, *supra* note 126.

137. *Id.*

138. Telephone Interview with Jud Hoffman, *supra* note 132.

account of their social status and the public interest in their doings.[139] According to Hoffman, this approach reflected Facebook's mission statement, which at that time was "Make the world more open and connected."[140] "Broadly, we interpreted 'open' to mean 'more content.' Yes, that's a bit of a 'free speech' perspective, but then we also had a concern with things like bullying and revenge porn," Hoffman recalled.[141] "But while trying to take down that bad content, we didn't want to make it impossible for people to criticize the president or a person in the news. It's important there's a public discussion around issues that affect people, and this is how we drew the line."[142]

In trying to resolve these dilemmas, Hoffman and his colleagues sought to "focus on the mission" of Facebook rather than adopt "wholesale . . . a kind of U.S. jurisprudence free-expression approach."[143] They quickly realized, however, that the mission had to be balanced against competing interests such as users' safety and the company's bottom line.[144] While Hoffman and Willner were at Facebook, the balance was often struck in favor of "leaving content up," but they were always searching for new ways to address concerns about harmful speech.[145] "We felt like Facebook was the most important platform for this kind of communication, and we felt like it was our responsibility to figure out an answer to this," said Hoffman.[146]

The policy required a way to determine if someone was a public figure. When a piece of content was flagged for bullying, Facebook told its moderators to use the news aggregator Google News.[147] If the person allegedly being bullied appeared in a Google News search, moderators would consider her a public figure—and the content would stay up.[148] Although some prominent people had blue "verification" checkmarks or "public figure" titles on their Facebook pages, these symbols were not actually part of the metric used to determine whether someone was a public figure under Facebook's Community Standards.[149] As Willner put it,

---

139. *Id.*
140. *Id.*
141. *Id.*
142. *Id.*
143. *Id.*
144. *Id.*
145. *Id.*
146. *Id.*
147. Telephone Interview with Dave Willner, *supra* note 132.
148. *Id.*
149. *Id.*

> The blue checkmark was totally separate from the "public figure" designation—an individual user getting a checkmark was much more arbitrary and came from a totally different team. Ultimately, it had no impact on how you were enforced against as a private or public figure—it just meant that you're one of the cool kids.[150]

Facebook's use of Google News to make public-figure determinations provided the platform with a tool that its moderators could use quickly and consistently. The ability to implement the underlying policy on such a mass scale was, of course, its virtue, at least in Facebook's eyes. But it was not without its drawbacks, as even Facebook's policymakers recognized. If anyone who appeared in a Google News search became a public figure, there was no way to know whether they had voluntarily entered the public eye. This issue often arose when people were caught up in terrible circumstances or publicly shamed in a way that went viral. As Willner put it, "you can think of them as involuntary public figures, but another way of saying it might be to think of them as sympathetic public figures."[151] Whether it was fair to apply the same bullying rules to these people was a question that Google's algorithm was simply unequipped to answer.

But not all involuntary public figures were plainly sympathetic. Consider, for example, the case of Casey Anthony, who became a household name after being accused of murdering her daughter in the "Social-Media Trial of the Century."[152] Or recall the furor surrounding *Rolling Stone*'s article about an alleged rape at the University of Virginia, a story that was later retracted and dubbed "a complete crock."[153] In both episodes, it was difficult for Facebook to identify the "sympathetic" parties caught up in the online firestorm that engulfed the platform, especially as the winds changed. With Casey Anthony, Willner recalled that Facebook felt torn between accepting the court's "not guilty" verdict and recognizing the hard reality that "everyone in America thinks she killed her kid."[154] Similarly, as Willner recounted:

---

150. *Id.*

151. *Id.*

152. *See* John Cloud, *How the Casey Anthony Murder Case Became the Social-Media Trial of the Century*, TIME (June 16, 2011), http://content.time.com/time/nation/article/0,8599,2077969,00.html [https://perma.cc/PB3Y-CLTU].

153. Erik Wemple, *Charlottesville Police Make Clear that Rolling Stone Story Is a Complete Crock*, WASH. POST (Mar. 23, 2015), https://www.washingtonpost.com/blogs/erik-wemple/wp/2015/03/23/charlottesville-police-make-clear-that-rolling-stone-story-is-a-complete-crock [https://perma.cc/94PA-532Q]; *see also* Ravi Somaiya, *Rolling Stone Article on Rape at University of Virginia Failed All Basics, Report Says*, N.Y. TIMES (Apr. 5, 2015), https://nyti.ms/1NMyThP [https://perma.cc/C5ZM-ZFRM].

154. Telephone Interview with Dave Willner, *supra* note 132.

> With the *Rolling Stone* story, when it starts, we believe the victims, and that people shouldn't say mean things about the victims, but then it turns out all that's not true—the victim there was actually the "bad" person. But where along in the journey of learning about that entire story do people's minds shift and what do you decide to protect?[155]

Despite the fact Google News could hardly provide the nuance to handle these edge cases, the company stood by its use as the best way to strike the balance between promoting free speech and remedying harmful speech.

In moderating content related to public figures, Facebook's policymakers began to blur the lines between social status and newsworthiness, just as judges have done when applying the First Amendment to privacy torts.[156] Willner reflected that "calling the exception [an exception for] 'public figures' was probably a mistake—a more accurate way of thinking about it is as a newsworthy person."[157] All that a Google News search could tell moderators was that someone's name had appeared in a news source—it could not reliably reveal the person's true social status, the reputability of the source, the veracity of the story, or the genesis of the controversy. This framework meant that Facebook ran into many of the same issues that have plagued courts in defining the boundaries of newsworthiness. As Willner recounted, he saw "newsworthiness" as representing a normative "post-hoc judgment that applies to the content as it's *supposed* to be—to be able to accurately assess it at the time literally calls for time travel."[158]

Facebook's engineers, while brilliant, had not realized that particular Sci-Fi dream, so the platform had to settle for a descriptive concept that included everything published in the "news." But this approach was not without its faults, and both Willner and Hoffman foresaw problems created by the erosion of traditional media, the rise of self-publishing, and the opportunities for mass amplification and virality offered by social media. "Now that everyone is their own newsroom, it's revealed that the emperor has no clothes," recounted Willner.[159] "We don't like how democratized reporting has gotten—everyone can be their own news station, and it's very upsetting to people."[160] Nevertheless, given the volume of content on Facebook and the subjectivity and unpredictability that would afflict case-

155. *Id.*
156. *See supra* Section I.B.
157. Telephone Interview with Dave Willner, *supra* note 132.
158. *Id.*
159. *Id.*
160. *Id.*

by-case newsworthiness determinations, the platform saw no viable alternative to this broad deference to an increasingly unprofessional media ecosystem. Moderators needed a tool to make quick and mechanical decisions, and a normative newsworthiness standard was too relative to measure consistently. As Hoffman remarked, "When we talk about a newsworthiness standard, what do we mean? Newsworthy to who and how many? If you don't establish a minimum number of people, then random gossip is newsworthy. How is somebody sitting in one of the [Facebook] operations places . . . going to decide that?"[161] For similar reasons, both Hoffman and Willner opposed building a general exception into the Community Standards to prevent removal of all "newsworthy" content. Facebook's approach to this issue would develop on a slightly different track.

B. NEWSWORTHY CONTENT

For most of Facebook's history, the platform made no exceptions for content that violated Community Standards but was newsworthy.[162] Overtly sexual, graphically violent, or "extremist" content would be taken down regardless of whether it had cultural or political significance.[163] This was a deliberate choice made by Hoffman and Willner, but the policy came under increasing pressure in recent years.[164]

Members of the policy team recall an incident in 2013 as a turning point toward the creation of an exception for newsworthy content.[165] In the wake of the Boston Marathon bombing, a graphic image of a man in a wheelchair began to circulate on Facebook.[166] The man was being wheeled away from the carnage, one leg ripped open below the knee to reveal a long, bloody bone.[167] What made this moderation question so fascinating was that there were three versions of the photograph.[168] One was cropped so that the leg was not visible.[169] A second was a wide-angle shot in which the leg was visible but less obvious.[170] The third, and most controversial, clearly showed the man's "insides on the outside"—the content-moderation team's

161. Telephone Interview with Jud Hoffman, Former Glob. Policy Manager, Facebook (Mar. 8, 2018).
162. *Id.*
163. Telephone Interview with Dave Willner, *supra* note 132.
164. *Id.*; Telephone Interview with Jud Hoffman, *supra* note 132.
165. Adler, *supra* note 129.
166. *Id.*
167. *Id.*
168. *Id.*
169. *Id.*
170. *Id.*

shorthand rule for when content was graphically violent.[171] Despite the fact that multiple media outlets had published all three photographs, Facebook removed any links to or images of the third version.[172] "Philosophically, if we were going to take the position that ['insides on the outside'] was our definition of gore and we didn't allow gore, then just because it happened in Boston didn't change that," remembers one of the team members on call that day.[173] Policy executives at Facebook disagreed, however, and reinstated all such posts on the grounds of newsworthiness.

For some members of the policy team, who had spent years trying to create administrable rules, the imposition of such an exception represented a radical departure from the company's commitment to procedural consistency. Some of their complaints echo the *Gertz* Court's rationale for reining in the plurality's rule in *Rosenbloom*.[174] In his opinion for the Court in *Gertz*, Justice Lewis Powell worried openly about allowing "judges to decide on an ad hoc basis which publications address issues of 'general or public interest' and which do not."[175] Many at Facebook worried similarly that "newsworthiness as a standard is extremely problematic: the question is really one of 'newsworthy to whom?' and the answer to that is based on ideas of culture and popularity."[176] The result, some feared, would be a mercurial exception that would, moreover, privilege American users' views on newsworthiness to the potential detriment of Facebook's users in other countries.[177]

Although there were other one-off exceptions made for incidents like the Boston Marathon photograph, Facebook's internal content-moderation policies continued to have no general exception for newsworthiness until September 2016, when a famous Norwegian author, Tom Egeland, posted a well-known historical picture to his Facebook page.[178] The photograph, "The Terror of War," depicts a nine-year-old Vietnamese girl naked in the street after a napalm attack, and for this reason the photo is often called "Napalm Girl."[179] In part because of its graphic nature, the photo was a pivotal piece

171. *Id.*
172. *Id.*
173. Telephone Interview with Anonymous, Former Member of Policy Team, Facebook (Aug. 28, 2018).
174. *See supra* notes 18–22 accompanying text.
175. Gertz v. Robert Welch, Inc., 418 U.S. 323, 346 (1974).
176. Telephone Interview with Anonymous, *supra* note 173.
177. *See* Adler, *supra* note 129.
178. *Id.*
179. *Id.*

of journalism during the Vietnam War.[180] Nonetheless, it violated Facebook's Community Standards.[181] Accordingly, Facebook removed the photo and suspended Egeland's account. Because of Egeland's stature, the takedown itself received news coverage. Espen Egil Hansen, the editor-in-chief of the Norwegian newspaper *Aftenposten*, published a "letter" to Facebook founder Mark Zuckerberg on *Aftenposten*'s front page calling for Facebook to take a stand against censorship. Hours later, Facebook's chief operating officer Sheryl Sandberg admitted that the company had made a mistake and promised that the rules would be rewritten to allow for posting of the photo.[182] Shortly thereafter, Facebook issued a press release underscoring the company's commitment to "allowing more items that people find newsworthy, significant, or important to the public interest—even if they might otherwise violate [its] standards."[183]

The "Terror of War" incident led Facebook to start looking more broadly at how it evaluated newsworthiness outside the context of cyberbullying. "After the 'Terror of War' controversy, we realized that we had to create new rules for imagery that we'd normally want to disallow, but for context reasons that policy doesn't work," said Peter Stern, head of Product Policy Stakeholder Engagement at Facebook.[184] According to Stern, this policy shift "led us to think about newsworthiness across the board."[185] He acknowledged that Facebook has two considerations when making newsworthiness determinations: "Safety of individuals on the one hand and voice on the other."[186] But what exactly does "voice" mean in this context? Here, again, Facebook has increasingly aligned itself with the courts' view of the relationship between free speech and self-governance. "When someone enters the public eye," Stern explained, "we want to allow a broader

180. *See* Kate Klonick, *Facebook Under Pressure*, SLATE (Sept. 12, 2016, 2:48 PM), http://www.slate.com/articles/technology/future_tense/2016/09/facebook_erred_by_taking_down_the_napalm_girl_photo_what_happens_next.html [https://perma.cc/P2WX-YZ5U].

181. The photo was likely removed because of the nudity, not because it was child pornography. *See* Kjetil Malkenes Hovland & Deepa Seetharaman, *Facebook Backs Down on Censoring 'Napalm Girl' Photo*, WALL ST. J. (Sept. 9, 2016, 3:07 PM), http://www.wsj.com/articles/norway-accuses-facebook-of-censorship-over-deleted-photo-of-napalm-girl-1473428032 [https://perma.cc/5GNF-DSKD].

182. *See* Claire Zillman, *Sheryl Sandberg Apologizes for Facebook's 'Napalm Girl' Incident*, TIME (Sept. 13, 2016), http://time.com/4489370/sheryl-sandberg-napalm-girl-apology [https://perma.cc/L8ME-Z637].

183. Joel Kaplan & Justin Osofsky, *Input From Community and Partners On Our Community Standards*, FACEBOOK NEWSROOM (Oct. 21, 2016), https://newsroom.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards [https://perma.cc/X48L-TYSD].

184. Telephone Interview with Peter Stern, Head of Prod. Policy Stakeholder Engagement, Facebook (Mar. 7, 2018).

185. *Id.*

186. *Id.*

scope of discussion."[187]

C. CONTEMPORARY APPROACHES

Over the last two years, Facebook's content-moderation policies have evolved to become somewhat less mechanical and more nuanced. For example, Facebook modified its rules on bullying and harassment of public figures in 2018. In place of its blanket rule that public figures could never enjoy protection from bullying, the platform adopted more context-sensitive standards to address some forms of harmful speech about prominent people. "Our new policy does not allow certain high-intensity attacks, like calls for death, directed at a certain public figure," members of the Facebook policy team reported on a recent call.[188] In the past, they explained, a statement such as "Kim Kardashian is a whore" would never be removed for bullying or harassment (whereas a statement calling a private individual a "whore" would be). But now, Facebook will remove some speech directed at public figures when it is posted on their own pages or accounts, depending on the severity of the language.[189] Details of how moderators will make these decisions are still vague, but it appears that the platform is beginning to draw lines based on both substance (whether the speech is particularly vicious) and directionality (whether the speech is targeted to reach particular public figures).[190]

Under this new regime, the platform has also tweaked its methods for determining someone's status as a public figure. Google News remains an important instrument in the moderator's toolkit—the platform will bestow the label on people mentioned in multiple news stories within a certain timeframe—but there are now other ways to qualify.[191] Regardless of what Google's algorithm regurgitates, Facebook now counts among its public figures people elected or assigned through a political process to a government position; people with hundreds of thousands of fans or followers on a social-media account; and people employed by a news organization or who speak publicly.[192]

Facebook's current policies on newsworthy content are somewhat

---

187. *Id.*

188. Televideo Interview with Idalia Gabrielow & Peter Stern, Policy Risk Team, Facebook (Aug. 15, 2018).

189. *Id.*

190. In this sense, Facebook and the court might again be following a similar path. *See supra* notes 100–01 and accompanying text.

191. Televideo Interview with Idalia Gabrielow & Peter Stern, *supra* note 188.

192. *Id.*

harder to pin down. Unlike the term "public figures," which Facebook still uses primarily for its bullying standards, "newsworthiness" is now a possible exception to *all* of the company's Community Standards.[193] And unlike the public-figure designations made in the bullying context, newsworthiness determinations do not rely on news aggregators and algorithms.[194] Instead, Facebook employees review claims about possible newsworthy content on a case-by-case basis.[195]

In deciding whether to keep up otherwise-removable content because of its newsworthiness, Facebook officials stress that they weigh the value of "voice" against the risk of harm.[196] Assessments of harm are informed by the nature as well as the substance of the objectionable content.[197] Hateful speech on its own, for instance, might be seen as less harmful than a direct call to violence.[198] Facebook officials maintain, however, that most of the newsworthiness decisions relate to nudity. Difficult decisions include what to do about nudity in public protests. "Just a few years ago, we took that down," stated David Caragliano, a policy manager at Facebook.[199] "But it's really important to leave this up consistent with our principles of voice. That's led to a policy change that's now at scale for the platform."[200] The non-hateful, nonviolent expressive conduct of public protesters, it seems, will today almost always be considered newsworthy and therefore will not be taken down, even if it runs afoul of other Community Standards.

Compared to the thousands of day-to-day decisions made by hordes of content moderators who compare content to rules, the "how" and "who" behind Facebook's newsworthiness determinations are more obscure. It is unclear, for example, how a question of possible newsworthiness climbs the Facebook policymaking ladder to become a new "policy change . . . at scale for the platform," let alone who makes that crucial call.[201] The lack of transparency and accountability gives little comfort to those who worry about the mercurial and subjective nature of newsworthiness determinations at Facebook.

This might be about to change. In November 2018, in "A Blueprint for

193. *Id.*
194. *Id.*
195. *Id.*
196. Telephone Interview with Peter Stern, *supra* note 184.
197. Televideo Interview with Ruchika Budhraja, David Caragliano, Idalia Gabrielow & Peter Stern, Policy Risk Team, Facebook (Oct. 4, 2018).
198. *Id.*
199. *Id.*
200. *Id.*
201. *See id.*

Content Governance and Enforcement," Mark Zuckerberg informed the public that he "increasingly [has] come to believe that Facebook should not make so many important decisions about free expression and safety on [its] own."[202] He announced that the platform would create an "Independent Governance and Oversight" committee to make decisions about the kinds of content users could post on the site.[203] Some have imagined this new body as a "Supreme Court" of Facebook.[204] Indeed, Zuckerberg himself used this analogy on an April 2018 podcast:

> You can imagine some sort of structure, almost like a Supreme Court, that is made up of independent folks who don't work for Facebook, who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world.[205]

The platform is now in the process of soliciting feedback on this idea and intends to establish the tribunal by the year's end.[206]

Whatever form Facebook's Supreme Court takes, it will surely face questions about the scope and application of Facebook's "doctrine" concerning public figures and newsworthiness. Actual courts, meanwhile, will continue to grapple with these concepts and may face similar questions as they confront challenges posed by tort claims arising from online speech. With this in mind, Part III now considers the lessons to be learned by comparing the public and private approaches to public figures and newsworthiness. The constitutional law created by the Old Governors greatly influenced the content moderation implemented by the New Governors, and the New Governors' experiences might now enlighten the Old Governors in turn.

III. FACEBOOK VERSUS *SULLIVAN*: LESSONS FOR COURTS AND PLATFORMS IN THE DIGITAL AGE

As we have seen, the two governance systems now used to adjudicate complaints about harmful speech—tort lawsuits in courts and content

---

202. Zuckerberg, *supra* note 123.

203. *See* Klonick & Kadri, *supra* note 5.

204. *Id.*; evelyn douek, *Facebook's New 'Supreme Court' Could Revolutionize Online Speech*, LAWFARE (Nov. 19, 2018, 3:09 PM), https://www.lawfareblog.com/facebooks-new-supreme-court-could-revolutionize-online-speech [https://perma.cc/PP5C-F668]; Kadri, *How Supreme a Court?*, *supra* note 5.

205. Ezra Klein, *Mark Zuckerberg on Facebook's Hardest Year, and What Comes Next*, VOX (Apr. 2, 2018, 6:00 AM), https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge [https://perma.cc/YTE5-JREW].

206. Zuckerberg, *supra* note 123.

moderation on platforms—share some similarities. Both have developed rules that seek to regulate harmful speech while protecting robust public discourse that is essential to self-governance. To strike this balance, courts and platforms have developed special rules that depend on the content of the speech and the parties involved in the dispute. In the courts, defamation law gives plaintiffs recourse for untruthful speech about them, but places a substantially higher burden on plaintiffs who are public figures.[207] On Facebook, an anti-bullying policy allows users to remove malicious speech about them, but users who are public figures can rarely avail themselves of this option.[208] In the courts, privacy and IIED law allow plaintiffs to hold defendants liable for certain privacy invasions or outrageous conduct, except when the underlying speech is deemed to be of legitimate public interest.[209] On Facebook, users can request that disturbing content like graphically violent or hateful speech be taken down, except when the content is of legitimate public interest.[210] Both judges and Facebook policymakers justify these rules in similar terms, citing the importance of protecting free speech and the legitimate public interest in discussion about people who are powerful, famous, or at the forefront of a particular controversy.[211]

The observation that First Amendment concepts like public figures and newsworthiness have wended their way into Facebook's content-moderation policies is interesting as a descriptive matter. It is part of a broader story about how American laws and norms have gained influence across the globe as these potent American companies have expanded their reach abroad.[212] Even though Facebook need not adhere to the First Amendment, its content-moderation policies were largely developed by American lawyers trained and acculturated in American free-speech norms, and it seems that this cultural background has affected their thinking.[213] By accurately describing some of Facebook's internal processes for moderating content in the way that it does, we can better understand Facebook's power inside and outside of the United States—and perhaps some of the external resistance to it, as

207. *See supra* Section I.A.

208. *See supra* Section II.A.

209. *See supra* Sections I.B–.C.

210. *See supra* Section II.B.

211. *Compare, e.g.*, N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964) (justifying heightened protection for speech about public figures because "debate on public issues should be uninhibited, robust, and wide-open"), *with* Telephone Interview with Peter Stern, *supra* note 184 (explaining that Facebook "want[s] to allow a broader scope of discussion" once "someone enters the public eye").

212. *See* Danielle Keats Citron, *What to Do About the Emerging Threat of Censorship Creep on the Internet*, CATO INST. (Nov. 28, 2017), https://www.cato.org/publications/policy-analysis/what-do-about-emerging-threat-censorship-creep-internet [https://perma.cc/ATU4-DCH6].

213. *See* Klonick, *supra* note 5, at 1621–22.

non-Americans bristle at the exportation of foreign values.[214]

But comparing these public and private governance systems provides more than mere descriptive interest: it also teaches important lessons about the challenges posed in the new speech ecosystem created by digital discourse today. As an initial matter, the comparison tests intuitions about why there are different standards concerning public figures—whether through the courts or private platforms—and how those standards relate to protections for newsworthy speech. With these rationales exposed, the comparison also reveals problems with how courts and platforms have implemented their standards for public figures and newsworthiness, especially in an era when online speech often influences public discourse. Finally, the comparison lays the groundwork to address some of those problems—a task briefly undertaken in this Article's conclusion.

A. THE RATIONALES BEHIND THE RULES

As this Article has revealed so far, courts and platforms have proffered various rationales for creating rules that protect newsworthy speech or hamstring public figures subjected to harmful speech. Unpacking these ideas tests intuitions about whether those rationales are both descriptively sound and normatively desirable. This scrutiny is particularly important today as the challenges posed by the digital age force courts to rethink old doctrines and lead the public to demand more from the private platforms that now govern speech online.

There have historically been three reasons why public figures have faced harsher standards than private figures when seeking recourse for harmful speech. The first is that public figures are supposedly "less vulnerable" because they have greater access to "channels of effective communication" to rebut the harmful speech.[215] This rationale—which the Court has also described as a public figure's greater ability to perform "self-help"—rests on an empirical postulate: that there is a meaningful difference between the abilities of public and private figures to engage in counter-speech that in some way redresses the harm.[216] This postulate explains, at

214. It is worth noting, however, that similar concepts exist in jurisprudence outside of the United States. For various international cases dealing with these examples, see generally Campbell v. MGN Ltd. [2004] 2 AC 457 (HL) (Eng.); Reynolds v. Times Newspapers [1999] 4 All ER 609 (HL) (Eng.); Lingens v. Austria, App. No. 9815/82, 8 Eur. H.R. Rep. 407 (1986); Tammer v. Estonia, App. No. 41205/98, 37 Eur. H.R. Rep. 43 (2001); Von Hannover v. Germany, 2004-VI Eur. Ct. HR 294.

215. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979).

216. *Id.*; *see also* Time, Inc. v. Hill, 385 U.S. 374, 391 (1967) (declining to base a constitutional

least in part, why the Court has adopted this rationale in the defamation context only. When the harm caused by speech stems from untruth that tarnishes someone's reputation, counter-speech can be an effective tool to rebut the lie and thereby address the injury.[217] But in the context of privacy or emotional harms caused by speech, counter-speech does little to ameliorate damage, which helps explain why a plaintiff's status as a public or private figure does little to change his or her legal rights when bringing privacy or IIED claims.[218] It might also explain why Facebook's policymakers never raise this rationale in explaining their public-figure rules related to cyberbullying, which again involves an array of speech harms that are not easily redressed through counter-speech.[219] Indeed, both offline and online, the harm caused by the types of speech that usually trigger privacy, IIED, and cyberbullying concerns might actually *increase* with counter-speech, either by amplifying the speech through more publicity or by forcing a victim to re-create or re-experience the harm in order to speak out against it.[220]

The second rationale for applying harsher rules to public figures is that they are "less deserving" of protection because they assume the risk of possible negative attention when they put themselves in the public eye.[221] The Court has explained that this is a "normative" rationale that depends on the idea of voluntariness—the harsher rule is justified because public figures, unlike private figures, have "voluntarily exposed themselves to increased risk of injury."[222] Once again, this reasoning appears crucial only in defamation jurisprudence where the Court has repeatedly cabined the scope of public-figure status by analyzing whether the plaintiff voluntarily "thrust" herself into the vortex of a public controversy.[223] (And while the Court flirted with the idea of an "involuntary" public figure, the justices have conspicuously left this character undefined and undiscovered in the forty years since he or she appeared hypothetically in *Gertz*.[224]) In privacy and

---

rule for *privacy* on "the distinction . . . between the relative opportunities of the public official and the private individual to rebut" harmful speech because defamatory speech creates a different type of harm).

217. *Wolston*, 443 U.S. at 164; *see also* Lyrissa Barnett Lidsky, *Defamation, Reputation, and the Myth of Community*, 71 WASH. L. REV. 1, 7 (1996).

218. *See supra* Sections I.B–.C.

219. *See supra* Section II.A.

220. Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 149–50 (2017) ("Efforts to remove hurtful material typically backfire by drawing additional attention to it, intensifying and prolonging the unwanted exposure.").

221. *Wolston*, 443 U.S. at 164.

222. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

223. *See, e.g.*, Hutchinson v. Proxmire, 443 U.S. 111, 135 (1979).

224. *Supra* note 41 and accompanying text.

IIED law, meanwhile, this rationale has no purchase: for the former, courts draw no doctrinal distinction between voluntary and involuntary public figures who bring claims for public disclosure of private facts;[225] for the latter, the Court in *Snyder* refused to give analytical weight to the plaintiff's status as a private figure involuntarily caught up in a newsworthy protest.[226] Facebook, too, has neglected to adopt this rationale, though it is unclear whether this is a decision based on principle or practicality.[227] Early on, some Facebook policymakers expressed concern about how the platform's algorithmic approach could not test for voluntariness, yet they still adopted that approach.[228]

The final reason to treat public figures differently rests on the idea that their prominence makes them subjects of "legitimate public concern." Under this rationale, which hews most closely to the self-governance theory of the First Amendment, the social status of public figures serves as a proxy for their newsworthiness, and public figures thus face harsher standards for the sake of cultivating robust public discourse.[229] This justification is most visible in the standards surrounding privacy law, which immunize disclosure of certain sensitive facts as matters of "legitimate public concern" only when they relate to a *public* figure.[230] Facebook's rules surrounding public figures have a similar flavor. Using Google News to make status determinations means that a person becomes a public figure on Facebook simply by appearing in a news story. This mechanism tells the platform nothing material about the person, aside from the fact that at least one news source decided that the person was newsworthy. This is, of course, a purely descriptive conception of newsworthiness—the person is newsworthy because they appear in the news—but it nonetheless tracks the proxy rationale to some extent.

This discussion reveals that these various rationales can take different forms: they can be grounded in descriptive claims about public figures, such as the notion that prominent people can more easily engage in effective self-help; or they can stem from normative considerations, including ideas of

---

225. *Compare* RESTATEMENT (SECOND) OF TORTS § 652D cmt. e (AM. LAW INST. 1977) (commenting on voluntary public figures), *with id.* § 652D cmt. f (commenting on involuntary public figures).

226. *See* Snyder v. Phelps, 562 U.S. 443, 474–75 (2011) (Alito, J., dissenting).

227. *See* Telephone Interview with Dave Willner, *supra* note 132.

228. *Id.*

229. *See, e.g.*, Gertz v. Robert Welch, Inc, 418 U.S. 323, 345 (1974) (explaining that, in defamation law, public figures must satisfy a heightened burden because of their "roles of especial prominence in the affairs of society").

230. RESTATEMENT (SECOND) OF TORTS § 652D cmt. d.

fairness, risk assumption, and legitimate scope of public curiosity. Ultimately, courts and platforms have used these descriptive and normative rationales to craft rules aimed at creating a speech ecosystem that preserves robust debate and protects against harmful speech. Whether they have done so effectively is another question—one that the following Section now attempts to answer.

B. JUDGING THE GOVERNORS, NEW AND OLD

Drawing out the comparisons between the public and private approaches to newsworthiness and public figures can teach valuable lessons about free speech in terms of both constitutional law and content moderation. This Section distills three insights that should enlighten both courts and platforms.[231] Although Facebook policymakers may have channeled doctrines created by courts, the platform's algorithmic approach fails to implement important normative protections baked into the courts' jurisprudence to justify harsher rules for public figures. Courts and platforms should also rethink their approaches to defining public figures in light of new phenomena created by social media, particularly given how virality alters ideas about voluntariness. Finally, platforms could assuage concerns raised by their ad hoc newsworthiness determinations by adopting structural changes to become more like the court system.

1. The Inaccuracies and Injustices of Algorithmic Authority

The first lesson learned from a comparative analysis stems from Facebook's approach to unearthing public figures. Facebook's use of Google News to determine whether a person is a public figure provides a vivid illustration of the problems raised when such definitions are outsourced purely to the media marketplace. Although the platform has suggested that this policy may be changing slightly, Facebook's method for ascertaining "public figure" status has traditionally turned on the presence or absence of an individual's name in news search results, which are effectively an averaging algorithm of media outlets' publication decisions. This runs straight into the threat of what Clay Shirky has called "algorithmic authority," insofar as "an unmanaged process of extracting value from diverse, untrustworthy sources" is treated as authoritative without any

231. This Section incorporates and builds upon material in Kate Klonick, *Facebook v.* Sullivan, KNIGHT FIRST AMEND. INST. (Oct. 1, 2018), https://knightcolumbia.org/content/facebook-v-sullivan [https://perma.cc/W9R8-7K9B].

human second-guessing or vouching for the validity of the outcome.[232]

As commentators have pointed out for over fifty years in a closely related context, if "newsworthiness" is defined solely in terms of news outlets' publication decisions, then granting a special legal privilege for newsworthy content is liable to swallow torts such as invasion of privacy. "The publisher has almost certainly published any given report because he judged it to be of interest to his audience, and believed that it would encourage them to purchase his publications in anticipation of more of the same," a student comment observed in 1963.[233] "A plaintiff in a privacy action would thus have lost almost before he started."[234] Partly for this reason, courts making these determinations have considered a range of factors[235] and, especially in recent years, have been unwilling to defer entirely to the media.[236] They have, in other words, refused to adopt a purely descriptive conception of newsworthiness that abdicates definitional responsibilities to the media.

The problems inherent in Facebook's approach are exacerbated by two features of the new speech ecosystem that platforms have themselves helped create. The first is that the nature of these platforms can, in some sense, create news, as when people become "Facebook Famous." For example, when a Facebook user killed a homeless man in Cleveland and then posted a video of the murder on Facebook, it was difficult to determine where the genesis

---

232. *A Speculative Post on the Idea of Algorithmic Authority*, CLAY SHIRKY (Nov. 15, 2009), http://www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority [https://perma.cc/64DR-P4TP]; *see also* James Grimmelmann, *The Platform is the Message*, 2 GEO. L. TECH. REV. 217, 217 (2018) (observing "the disturbing, demand-driven dynamics of the Internet today, where any desire, no matter how perverse or inarticulate, can be catered to by the invisible hand of an algorithmic media ecosystem that has no conscious idea what it is doing").

233. Comment, *The Right of Privacy: Normative-Descriptive Confusion in the Defense of Newsworthiness*, 30 U. CHI. L. REV. 722, 725 (1963).

234. *Id.*

235. *See, e.g.*, Snyder v. Phelps, 562 U.S. 443, 453 (2011) ("Deciding whether speech is of public or private concern requires us to examine the content, form, and context of that speech, as revealed by the whole record." (internal quotation marks omitted) (citation omitted)).

236. *See, e.g.*, Amy Gajda, *Judging Journalism: The Turn Toward Privacy and Judicial Regulation of the Press*, 97 CALIF. L. REV. 1039, 1041–42 (2009) (explaining that some courts have become less deferential to the media in determining newsworthiness, perhaps on account of "growing anxiety about the loss of personal privacy in contemporary society" or "declining public respect for journalism"); Sydney Ember, *Gawker and Hulk Hogan Reach $31 Million Settlement*, N.Y. TIMES (Nov. 2, 2016), https://www.nytimes.com/2016/11/03/business/media/gawker-hulk-hogan-settlement.html [https://perma.cc/TG5X-2W62] (describing the groundbreaking jury verdict that awarded former professional wrestler Hulk Hogan $140 million after the gossip news site Gawker.com published a sex tape featuring him). *But cf.* Carroll, *supra* note 95, at 77–81 (discussing cases in which the courts have "largely left the role [of determining what is newsworthy or of legitimate public interest] to the press").

of "publicity" begins.[237] As Hoffman described, "the problem with using Google News to determine public figure is that sometimes . . . the source of the Google News result would be Facebook."[238] This feedback loop undermines one potential virtue of delegating the public-figure determination to an outside algorithm: the notion that doing so would provide some legitimacy because it would defer to neutral sources external to the platform. There was never an official solution during Hoffman's tenure on how to respond to the circular moment of a person becoming a public figure because of their actions on the platform.[239]

The second—and related—problem is that Google News is a news aggregator, not a newsroom or newspaper. Courts have justified harsher standards for public figures because speech about them tends to be newsworthy.[240] Facebook's approach partially adopts this proxy rationale because a person's appearance in the news makes them a public figure.[241] But whereas judicial deference to the press has historically rested upon trust in the press as an institution, it is difficult to ascribe the same wisdom to Google's algorithm. Facebook's tactic of defining "newsworthy people" using not *the press* but *a news algorithm* raises concerns because someone can easily become "newsworthy" without "news judgment." In the new speech ecosystem brought about in part by platforms like Facebook, people can be thrust into the public sphere and bypass the gatekeeping function of the traditional press.[242] In short, there are no editorial desks at Google

237. Jonah Engel Bromwich, *Cleveland Police Seek Suspect After a Killing Seen on Facebook*, N.Y. TIMES (Apr. 16, 2017), https://www.nytimes.com/2017/04/16/us/facebook-live-shooting.html [https://perma.cc/Z7JC-FW3D].

238. Telephone Interview with Jud Hoffman, *supra* note 132.

239. *Id.* Professor Enrique Armijo has argued that First Amendment law would provide a "useful heuristic for content moderation" in this area because a "defamation defendant cannot cause a plaintiff to become a public figure by dint of the statements that gave rise to the claim." Enrique Armijo, *Meet the New Governors, Same as the Old Governors*, KNIGHT FIRST AMEND. INST., (Oct. 30, 2018), https://knightcolumbia.org/content/meet-new-governors-same-old-governors [https://perma.cc/5QPY-E45T]; *see also* Wells v. Liddy, 186 F.3d 505, 511, 541 (4th Cir. 1999) (explaining that the relevant controversy "must have existed prior to the publication of the defamatory statement"). Armijo argues that Facebook could channel "[g]ood old-fashioned First Amendment law" to set a new rule: "if the results of [the Google News] search include only stories about the complained-of bullying itself, then the victim of the bullying is a private person" and could still rely on the platform's protections against cyberbullying. Armijo, *supra*. Facebook might complain that this task will be hard for content moderators to handle at scale, but Armijo is correct that it might provide a fairer system than one that defers entirely to Google's algorithm.

240. *See, e.g.*, Rand v. Hearst Corp., 298 N.Y.S.2d 405, 411 (App. Div. 1969) (explaining that privacy invasions may be justified if the plaintiff "has achieved the position of a 'public figure' and thus became newsworthy" (citation omitted)).

241. Telephone Interview with Dave Willner, *supra* note 132.

242. *See infra* Sections III.B.2–.3.

News.[243]

All of this creates a kind of inverse Goldilocks principle whereby Facebook ends up removing too much benign speech *and* preserving too much harmful speech. As to the former, recall that Facebook's cyberbullying policies traditionally meant that any user could have offensive speech about them removed so long as they were not a public figure, and conversely all speech about public figures would stay up.[244] A purely descriptive approach to identifying public figures means that important or influential people slip through the cracks and remain "private" figures. News aggregators may struggle to capture localized power or notoriety in smaller communities—an issue made more problematic by the very nature of social media, which has enabled virtual communities to develop their own distinctive cultures and social structures.[245] Provocative content flagged in these communities may not seem to involve any "public figures" when judged against a global Google News search and therefore may be removed even if it involves a matter of intense interest within that community.

Facebook's algorithmic approach also prioritizes scalability over accuracy. Even by their own admission, the platform's policymakers knew that it could not ascertain whether someone was being cyberbullied by looking only at the flagged content.[246] The use of Google News was thus a haphazard way to ensure that prominent people could not use a powerful and blunt tool to remove critical speech about them, while still allowing private figures to remove potentially verboten content with a simple complaint.[247] According to the many different Facebook employees with whom we have spoken over the years, the vast majority of content that gets flagged for moderators is not speech that actually violates the platform's rules but rather speech that certain users simply do not like.[248] Making important distinctions based on an unnuanced mechanism like Google News will inevitably lead to false positives.

At the other end of the spectrum, Facebook's public-figure determinations can preserve too much harmful speech. As courts know all

---

243. *See* Amy Gajda, *Newsworthiness and the Search for Norms*, KNIGHT FIRST AMEND. INST. (Oct. 30, 2018), https://knightcolumbia.org/content/newsworthiness-and-search-norms [https://perma.cc/Y7US-2NBR].

244. *See supra* Section II.A.

245. *See* BOYD, *supra* note 8, at 7–14.

246. Telephone Interview with Dave Willner, *supra* note 132.

247. Telephone Interview with Jud Hoffman, *supra* note 132.

248. Telephone Interview with Dave and Charlotte Willner, *supra* note 126; Telephone Interview with Jud Hoffman, *supra* note 126.

too well,[249] deferring to the public's curiosity can sometimes lead to unjust results—and Facebook's approach effectively implements such blind deference. It does so in two ways, all of which can be understood as being part of a larger problem with algorithmic authority: unlike in the tort system, there is no "normative backstop" to prevent people facing unjustly harsh rules. First, as we discussed already, courts handling privacy claims have increasingly refrained from allowing the media to define newsworthiness by insisting that a disclosed fact is a matter of "legitimate" public concern.[250] The idea of legitimacy in this context allows courts and juries to consider community mores in defining newsworthiness as a way to prevent "morbid and sensational prying into private lives for its own sake."[251] Facebook's approach to defining public figures contains no such limitation; indeed, it is tough to imagine how a news algorithm could ever engage in such a complex and context-dependent inquiry. As a result, a Facebook user might be denied protection against cyberbullying even if he or she appeared in a news story (and thus became a "public figure" on the platform) that violated notions of "common decency."[252] This might strike many as an unjust result.

Facebook's approach lacks a second normative backstop that courts have built into the doctrine: the voluntariness requirement from defamation law. As we have seen, courts use the voluntariness rationale to limit harsher rules to plaintiffs who are "less deserving" of protection against defamation under the theory that they have assumed the risk of injury by *voluntarily* entering the public eye.[253] A Google News search cannot provide any such limitation. This failing is all the more concerning because involuntary public figures—once considered "exceedingly rare" by the Court[254]—are increasingly common in the online realm.[255] Countless stories exist of relatively unknown individuals being filmed or photographed and then finding themselves subject to widespread online shaming and related news coverage.[256] Should such an individual report any particularly offensive posts to Facebook for violating the company's cyberbullying rules, the Google News search results would indicate that the individual is a public figure and—at least until recently—the posts would stay up. Google News

249. *See* RESTATEMENT (SECOND) OF TORTS § 652D cmt. h (AM. LAW INST. 1977).

250. *See id.*

251. *Id.*

252. *See id.*

253. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979) (quoting Gertz v. Robert Welch Inc., 418 U.S. 323, 345 (1974)).

254. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

255. *See infra* Section III.B.2.

256. *See* Klonick, *supra* note 8, at 1044–50.

is unequipped to distinguish between situations in which people have voluntarily "thrust themselves to the forefront of particular public controversies"[257] and situations in which someone has been catapulted into internet stardom through little or no action of their own.[258] Concern about inequitable treatment for targets of harmful speech is precisely what led the Court in cases like *Firestone* to limit the class of individuals who would face the harsher defamation rules reserved for public figures, but Facebook has yet to respond to similar fears on its platform.[259]

2. Voluntariness and Sympathy in the Age of Virality

To be charitable to Facebook, the platform's inattention to the constitutional concept of voluntariness might reflect a skepticism about its normative desirability in the digital age. The Court has insisted that those who voluntarily enter the public arena are "less deserving" of protection because they assume the risk of negative attention,[260] but this premise has been undermined by certain features of the new speech ecosystem brought about by platforms like Facebook. For this reason, both courts and platforms should reform—and maybe even replace—their use of voluntariness as a bellwether to decide which people deserve harsher rules in disputes about harmful speech as a matter of constitutional law or content moderation.

For purposes of defamation claims, *Gertz*'s taxonomy splits the world into four types of people: general public figures, who are in "positions of such persuasive power and influence that they are deemed public figures for all purposes"[261]; limited-purpose public figures, who "thrust themselves to the forefront of particular public controversies in order to influence the

257. *Gertz*, 418 U.S. at 345.

258. Not only does Facebook's approach fail to test for voluntariness, it also features no way to enforce the "public controversy" limitation that plays an important role in defamation law. *See* Silvester v. Am. Broad. Cos., 839 F.2d 1491, 1494 (11th Cir. 1988) (explaining that "a public controversy must be more than merely newsworthy").

259. Perhaps aware of the inequities built into its approach, Facebook has attempted one trick to limit the scope of its public-figure designation: it limited the timeframe in which someone had to appear in the news to qualify as a public figure. This approach has echoes of the courts' treatment of "limited-purpose" public figures, who gain the status because of their voluntary involvement in a particular public controversy, but it is fundamentally different from the normative considerations that animate the doctrine. Indeed, Facebook's temporal approach creates censorship risks because important speech could be removed simply because the subject of the speech has not *recently* been in the news. *Cf.* RESTATEMENT (SECOND) OF TORTS § 652D cmt. k (AM. LAW INST. 1977) ("The fact that there has been a lapse of time, even of considerable length, since the event that has made the plaintiff a public figure, does not of itself defeat the authority to give him publicity or to renew publicity when it has formerly been given.").

260. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979) (quoting Gertz v. Robert Welch, Inc. 418 U.S. 323, 345 (1974)).

261. *Gertz*, 418 U.S. at 345.

resolution of the issues involved;"[262] involuntary public figures, who become famous "through no purposeful action of [their] own"[263]; and private figures, who are "more vulnerable to injury" and thus "more deserving of recovery."[264] While the first two groups "invite attention and comment" through their own purposeful actions, the last two do not—and, as a result, the Court has not applied the harsher constitutional rules to involuntary public figures or private figures in defamation actions.[265]

The same cannot be said for privacy law, which differs from defamation law in two important respects. First, privacy law appears to define public figures more broadly. The Restatement squashes the general- and limited-purpose public figure into one category—the "voluntary public figure"—and defines a public figure as anyone who "voluntarily places [oneself] in the public eye, by engaging in public activities, or by assuming a prominent role in institutions or activities having general economic, cultural, social or similar public interest, or by submitting [oneself] or [one's] work for public judgment."[266] This lacks the limitation that the Court has crafted in defamation law that limited-purpose public figures must embroil themselves in a "public controversy," which the Court tellingly refused to equate with "all controversies of interest to the public" because doing so would reinstate the *Rosenbloom* rule that disavowed the significance of a plaintiff's social status.[267] In other words, the "public controversy" requirement serves to limit the class of public figures in a way that has no analog in privacy law, where you can become a public figure by involving yourself in activities that have "public interest."[268]

The second difference between defamation and privacy law is that voluntary and involuntary public figures are treated identically in privacy law, whereas the Court has yet to clarify the constitutional consequence of being an involuntary public figure in defamation law. It seems, however, that the logic of the Court's post-*Gertz* jurisprudence could not support equal treatment for involuntary public figures who bring defamation claims. Decisions like *Firestone*, *Wolston*, and *Proxmire* all place great weight on

262. *Id.*
263. *Id.*
264. *Id.*
265. *See id.*
266. RESTATEMENT (SECOND) OF TORTS § 652D cmt. e (AM. LAW INST. 1977).
267. Time, Inc. v. Firestone, 424 U.S. 448, 454 (1976); *see also* Silvester v. Am. Broad. Cos., 839 F.2d 1491, 1494 (11th Cir. 1988).
268. RESTATEMENT (SECOND) OF TORTS § 652D cmt. e.

the idea of voluntariness,[269] which the Court has called the "normative consideration" that justifies the harsher rules that public figures face under defamation law.[270] In *Wolston*, for example, the Court refused to apply the public-figure label to the grand-jury witness who "was dragged unwillingly into the controversy."[271] This means that Facebook's rules for public figures hew closer to privacy law because the platform also draws no distinction based on voluntariness.[272]

To understand why the normative salience of voluntariness might have shifted with the rise of platforms like Facebook, it is helpful to run through a few real-world examples to test intuitions. With each example, consider whether applying a harsher rule (in the courts or on platforms) seems fair based on a voluntary-involuntary distinction, and indeed whether that distinction has any influence on popular or legal intuitions. This allows courts and platforms to possibly reform the idea of voluntariness—and indeed decide whether they should replace it with something else entirely.

In November 2014, a Twitter user posted a photo of a Target employee bagging items behind a cashier.[273] The employee was wearing the nametag "Alex."[274] In one day, the tweet gained over one thousand retweets and two thousand favorites.[275] In two days, the "#AlexFromTarget" hashtag had over a million Twitter hits and the phrase "Alex From Target" racked up more than 200,000 Google searches.[276] Before long, Twitter users managed to identify "Alex" and unearth his Twitter account, which quickly amassed over 250,000 followers.[277] Alex appeared on the *Ellen* talk show two days later. Death threats, denigrating posts, and "fabricate[d] stories" about him soon followed.[278] It is hard to argue that Alex from Target, a "global celebrity" with hundreds of thousands of social-media followers,[279] is merely a private figure. Similarly, it is hard to argue that he is a voluntary public figure who

269. *See* Hutchinson v. Proxmire, 443 U.S. 111, 135 (1979); Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979); *Firestone*, 424 U.S. at 454.

270. *Wolston*, 443 U.S. at 164.

271. *Id.* at 166.

272. Telephone Interview with Dave Willner, *supra* note 132.

273. *Alex from Target / #AlexFromTarget*, KNOW YOUR MEME, https://knowyourmeme.com/memes/alex-from-target-alexfromtarget [https://perma.cc/ASB5-GMAK].

274. *Id.*

275. *Id.*

276. *Id.*

277. *Id.*

278. Nick Bilton, *Alex from Target: The Other Side of Fame*, N.Y. TIMES (Nov. 12, 2014), https://www.nytimes.com/2014/11/13/style/alex-from-target-the-other-side-of-fame.html [https://perma.cc/J62U-QNXF].

279. *Id.*

thrust himself into the vortex of a public controversy by bagging groceries at his part-time job.[280] At most, then, he is the elusive involuntary public figure that the Court believed was "exceedingly rare,"[281] though his fleeting internet stardom seems different from the public's interest in the two involuntary public figures mentioned in the Restatement—perpetrators and victims of crimes.[282]

Unlike Alex from Target, some people play a more active role in triggering the public's attention by posting content online that then goes viral. Consider the example of Justine Sacco, who tweeted before boarding a flight to Cape Town: "Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!"[283] The tweet went viral, and people all around the world began following the "#HasJustineLandedYet" hashtag to track her progress as they discussed what she had said.[284] When she landed in Cape Town eleven hours later, she discovered that she was "the No. 1 worldwide trend on Twitter" and had tens of thousands of responses to her tweet.[285] The story was picked up by several major media outlets,[286] and Sacco soon lost her job.[287] As with Alex from Target, it seems strange to think of Sacco as a private figure given her sudden worldwide fame. But whether to dub her a "voluntary" or "involuntary" public figure is far from clear. It seems like a stretch to say that, under defamation law, she voluntarily "thrust [herself] to the forefront of particular public controversies in order to influence the resolution of the issues involved,"[288] but she might qualify as a voluntary public figure under privacy law's more permissive standard.[289] Any designation is further complicated by the fact that the public furor on social media surrounding Sacco's tweet arguably created a newsworthy event in its own right, even if the underlying events did not.

Sacco, at least, was an adult who presumably had some sense of the possible ramifications of her actions, even if she could not have predicted their scale or intensity. Matters become more complicated, however, when

280. *See* Hutchinson v. Proxmire, 443 U.S. 111, 135 (1979).

281. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

282. RESTATEMENT (SECOND) OF TORTS § 652D cmt. f (AM. LAW INST. 1977).

283. Klonick, *supra* note 8, at 1048–49 (citation omitted).

284. Jon Ronson, *How One Stupid Tweet Blew Up Justine Sacco's Life*, N.Y. TIMES MAG. (Feb. 12, 2015), https://www.nytimes.com/2015/02/15/magazine/how-one-stupid-tweet-ruined-justine-saccos-life.html [https://perma.cc/3CD3-MY7E].

285. *Id.*

286. Klonick, *supra* note 8, 1048–49.

287. Ronson, *supra* note 284.

288. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

289. RESTATEMENT (SECOND) OF TORTS § 652D cmt. e (AM. LAW INST. 1977).

children become internet sensations. This can be so when they lack the maturity to grasp the consequences of their actions, as one might think occurred recently with students from Covington Catholic whose interaction with a Native American activist became a huge story after videos of the incident went viral on social media.[290] As one journalist quipped in the aftermath, "The thing to remember about the dumbass teens of Covington Catholic is that while they are dumbasses, they are also teens."[291] Even though the students voluntarily took part in a public march in the middle of the nation's capital while they knew people were filming events on their phones, to say that they assumed the risk of such colossal negative attention—in the way courts speak of public figures in defamation law—seems farfetched.[292]

Perhaps more troubling are those children who are "thrust" into the limelight by their parents. Six-year-old Adalia Rose Williams, for example, gathered nearly six million Facebook fans after her mother set up a page about her rare and fatal condition that made Williams age much faster than normal.[293] The Facebook page was a source of widespread support and empathy, but it also drew vile abuse and even a hoax story that Williams had died.[294] Williams's mother admitted that she was trying to raise awareness about her daughter's condition by placing her in the public eye.[295] Williams was, in a sense, "thrust" into a public forum created intentionally to provoke widespread engagement and discussion, but Williams herself did nothing to encourage her fame. Nonetheless, whether she voluntarily or involuntarily garnered public attention, it is hard to think of someone tracked by millions

290. Mike Pesca, *Covington Boys: The Difference Between Jerks and Monsters*, SLATE (Jan. 24, 2019, 7:16 PM), https://slate.com/news-and-politics/2019/01/covington-catholic-the-scandal-that-isnt-a-scandal.html [https://perma.cc/RSM2-UR6P].

291. *Id.*

292. *See* Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979). It is true, of course, that Covington students might not qualify as defamation public figures in the courts, despite their online fame. *See* Eugene Volokh, *Libel Law and the Covington Boys*, VOLOKH CONSPIRACY (Jan. 24, 2019, 1:01 PM), https://reason.com/volokh/2019/01/24/libel-law-and-the-covington-boys [https://perma.cc/TH9S-ZGWJ] (arguing that the students are still private figures because "they weren't famous or influential before this event" and that "just showing up at a rally would [not] qualify" as "voluntarily entering some particular debate" to make them limited-purpose public figures).

293. Simon Tomlinson, *Six-Year-Old Girl with Body of an Old Woman... Who Sings Vanilla Ice, Dances Gangnam Style and Has Own Fan Club: Adalia Rose Suffers Rare Premature Aging Condition*, DAILY MAIL (Feb. 25, 2013, 7:06 AM), https://www.dailymail.co.uk/news/article-2284110/Adalia-Rose-The-year-old-girl-body-old-woman-progeria.html [https://perma.cc/C3FH-ZCGY].

294. *Id.*

295. Jonathan Weiss, *Girl with Rare Genetic Disorder Bullied Online*, MED. DAILY (Feb. 26, 2013, 3:24 PM), https://www.medicaldaily.com/girl-rare-genetic-disorder-bullied-online-244525 [https://perma.cc/D9CS-UNZL].

of people as a private figure.

Finally, there are people who have chosen to be in the spotlight but who might nonetheless seem deserving of protection against certain types of online reactions. Take even the most obvious public figures, like actress and comedienne Leslie Jones, who understandably spend much of their time trying to gain public attention. Jones was inundated with racist and sexist comments on Twitter after she starred in the all-female *Ghostbusters* remake.[296] In the wake of the abuse, Jones tweeted "I feel like I'm in a personal hell. I didn't do anything to deserve this."[297] There is no doubt that Jones's fame makes her a public figure under defamation and privacy law. But whether—as a normative matter—she deserves the harsher cyberbullying rules that accompany public-figure status on Facebook is a far harder question.

What can these anecdotes teach courts and platforms as they handle disputes in the digital age? The first lesson is that the voluntary-involuntary distinction needs reform if it is to retain normative appeal as a barometer for how "deserving" a person is of harsher rules. A person's viral internet fame might make it difficult to conceive of her as a private figure, but it might also strike us as unfair if that person's decision to post on social media suffices to strip that person of protection under tort law or content-moderation rules. This intuition might stem from at least three points about life in the digital age. First, even if a person intentionally posts online, it is not safe to presume that he or she was intending to provoke an online firestorm and suddenly become famous. Second, we might question whether the person truly assumed the risk of a viral reaction and all of the baggage that can accompany it.[298] And third, it is no longer so "rare" to become involuntarily famous with little or no "purposeful action of [his or her] own."[299] These three points bring to mind a person walking through a forest when the heel of her shoe sets off a spark that creates a massive forest fire. He or she may have voluntarily taken actions in the world that led to a catastrophic incident, but it feels wrong to say that he or she assumed that risk or deserves to be

296. Anna Silman, *A Timeline of Leslie Jones's Horrific Online Abuse*, THE CUT (Aug. 24, 2016), https://www.thecut.com/2016/08/a-timeline-of-leslie-joness-horrific-online-abuse.html [https://perma.cc/2ACG-6STU].

297. *Id.*

298. *Cf.* Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974) (explaining that public figures "invite attention and comment" and thus have "voluntarily exposed themselves to increased risk of injury from defamatory falsehood concerning them").

299. *See supra* notes 257–81 and accompanying text.

punished for arson.[300]

This might mean that, in defamation law, courts should narrowly define public figures to prevent the exception from swallowing the rule. If, as one court has remarked, a public figure is "anyone who is famous or infamous because of who he is or what he has done,"[301] there will be far too many public figures in this world. Courts should vigilantly apply the limitations placed on public-figure status in defamation law, insisting that people have voluntarily embroiled themselves in a "particular public controversy" and not simply been swept up in events that suddenly become "of interest to the public."[302] At the very least, the Supreme Court should bring clarity to the constitutional significance of being an *involuntary* public figure. The Court's post-*Gertz* jurisprudence suggests that such people do not deserve harsher defamation rules, but the time has come for certainty now that they are no longer "exceedingly rare."

To reform privacy law, courts would need to both narrow the definition of a voluntary public figure and distinguish between voluntary and involuntary public figures. Under current privacy doctrine, a vast range of people could be voluntary public figures: posting a Tweet, photo, blog post, or status update would all seem to qualify as having "submitted [oneself] or [one's] work for public judgment" so as to make every human who has ever had a social-media account a voluntary public figure.[303] Even those who refrain from social media but get mentioned on it by "engaging in public activities" might qualify.[304] If not a voluntary public figure, such hapless individuals might be involuntary public figures, who are subject to identical treatment under privacy law. This lack of distinction is particularly troublesome because, as we have seen, platforms themselves have created an environment in which virality is a feature of their products and involuntary public figures are everywhere.[305]

---

300. To any graduate of an American law school, such an example might call to mind the classic torts case of *Palsgraf v. Long Island Railroad Co.*, 162 N.E. 99, 99 (N.Y. 1928), in which the plaintiff, Helen Palsgraf, was waiting to board a train when another passenger stumbled while running to catch the train, dropping a package of fireworks, which subsequently exploded and caused a large scale on the platform to hit her. The case is typically taught to introduce the idea of foreseeability and causation as potential limiting factors on tort liability. Foreseeability might provide insights for why we might not want to *increase* the burden on the plaintiff in certain instances.

301. Cepeda v. Cowles Magazines & Broad., Inc., 392 F.2d 417, 419 (9th Cir. 1968).

302. Time, Inc. v. Firestone, 424.U.S. 448, 453–54 (1976).

303. *See* RESTATEMENT (SECOND) OF TORTS § 652D cmt. e (AM. LAW INST. 1977).

304. *See id.*

305. *Cf.* James Grimmelmann, *Saving Facebook*, 94 IOWA L. REV. 1137, 1137 (2009) (observing that "Facebook offers a socially compelling platform that also facilitates peer-to-peer privacy violations" in part because "people use Facebook with the goal of sharing information about themselves").

Given the challenges posed by the digital age, perhaps the better approach is to replace voluntariness altogether as a normative measurement. Now that people voluntarily engage in all sorts of activities that unexpectedly spawn virality and fame, one might question whether voluntariness is doing the work that it once did to assess who is "less deserving" of protection against harmful speech. Put differently, has the ease with which everyone can now publish and amplify their speech changed the social (and perhaps legal) significance of voluntary engagement in public debate? Moreover, does the newfound prevalence of involuntary public figures undermine distinctions based on voluntariness now that it is no longer "exceedingly rare" to be thrust into the limelight against your will?

Platforms like Facebook have created an environment that reveals problems with penalizing people for "thrusting" themselves into a "public controversy." To apply harsher rules to people who speak out risks chilling valuable expression. This is particularly concerning when targets of harmful speech might surrender certain protections as the price of publicly responding to online abuse—a dynamic that risks "blaming the victim," heightening their vulnerability, and worsening the harm they might suffer.[306]

These concerns might explain why platforms like Facebook eschew considerations of voluntariness and instead reach for other normative concepts to judge what is "fair" in their systems of private governance. To borrow the language of one Facebook policymaker, what might matter is whether a person is "sympathetic." The sympathetic public figure is often involuntary, but not always. As Willner put it, "you can think of them as involuntary public figures, but another way of saying it might be to think of them as sympathetic public figures"[307] because they most frequently came up with people who were caught up in terrible circumstances or had been publicly shamed and that shaming had gone viral.

It is unclear whether Facebook ever formalized this concept in any actual rules, but it is nonetheless useful to consider how it might have influenced the platform's line-drawing in this context. Rather than using an empirical fact as a basis for a normative conclusion—such as determining

306. For discussion of this general problem online and in other areas of law, see DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 77–78 (2014); Heidi M. Hurd, *Blaming the Victim: A Response to the Proposal That Criminal Law Recognize a General Defense of Contributory Responsibility*, 8 BUFF. CRIM. L. REV. 503, 510 (2005); Josephine Ross, *Blaming the Victim: 'Consent' Within the Fourth Amendment and Rape Law*, 26 HARV. J. ON RACIAL & ETHNIC JUST. 1, 3 (2010); JoAnne Sweeny, *Gendered Violence and Victim-Blaming: The Law's Troubling Response to Cyber-Harassment and Revenge Pornography*, 8 INT'L J. TECHNOETHICS 18, 23 (2017).

307. Telephone Interview with Dave Willner, *supra* note 132.

whether somebody voluntarily invited public attention as a way to conclude that they are "less deserving" of protection—the description of someone as "sympathetic" skips straight to an opaque normative analysis about whether applying the harsh rule is fair under the circumstances. This makes it a tough standard to administer, particularly on a mass scale where such subjectivity can breed inconsistency. But the same can be said for legal standards that rest on community mores—such as the inquiry into whether a matter is of "legitimate" public concern in determining its newsworthiness. Perhaps, then, the "sympathetic" public figure is not so different from the person who, under privacy or IIED law, must face harsher rules for the sake of ensuring robust public discourse.

We might begin to give some shape to the concept of the sympathetic public figure by analyzing how it might fare better than the notion of voluntariness in the digital age. As we have seen, voluntariness can still influence the normative calculus in certain circumstances: Alex from Target might eventually become a public figure following his online stardom, but it feels odd to treat him identically to someone like Donald Trump given that Alex did nothing (aside from being handsome while bagging groceries!) to encourage his fame.[308] But we might also worry about how fair it is for courts and platforms to apply harsher rules to the likes of Justine Sacco, the Covington Catholic students, and Adalia Rose even though all of them, in some sense, became famous through voluntary actions taken by themselves or their legal guardians.[309] What ties these three anecdotes together is that it seems unlikely that the people involved foresaw that their actions would spur such a significant reaction and transform them into public figures, nor would a reasonable person have foreseen it either. It therefore seems unfair to say—as we might in defamation law—that they are "less deserving" of protection because they assumed the risk of becoming targets of harmful speech due to their own voluntary actions.[310]

If Facebook and other platforms are serious about replacing the concept of voluntariness with something like "sympathy," they cannot rely on algorithms and automation to do the job. These judgments are too nuanced and depend on understanding context and complicated social facts, so mechanical application of the standard will fail. Google News, then, must be replaced or supplemented by other tools that, at least at some point in the process, incorporate human review. Without it, platforms will run headlong

308. Bilton, *supra* note 278.
309. *See supra* notes 267–79 and accompanying text.
310. Wolston v. Reader's Digest Ass'n, 443 U.S. 157, 164 (1979).

into the same problems discussed in the previous Section, whereby algorithmic authority removes any normative backstop and implements a purely descriptive approach. Still, this case-by-case approach is not without its drawbacks, as we will now see.

3. The Perils of Ad Hoc Exceptions

Merely including humans in the mix is not necessarily a solution to the problems posed by algorithmic authority and the new speech ecosystem that platforms have helped create. Facebook's experiences implementing its ad hoc newsworthiness exception are proof of the perils of case-by-case adjudication in this area. Given courts' parallel experiences, this should come as no surprise. Indeed, similar concerns animated the Supreme Court's decision to overrule *Rosenbloom* in *Gertz*, in which Justice Lewis Powell warned of the dangers of "forcing state and federal judges to decide on an ad hoc basis which publications address issues of 'general or public interest' and which do not—to determine . . . 'what information is relevant to self-government.'"[311] If the Court "doubt[ed] the wisdom of committing this task to the conscience of judges" in the court system,[312] we might also worry about Facebook executives acting similarly behind closed doors.

As things stand, Facebook and other platforms seem to deviate from their rules if a high-ranking policymaker at the company intervenes and determines that a piece of content is newsworthy. This overarching exception applies to all types of content on Facebook. In the context of Napalm Girl, for instance, Sheryl Sandberg intervened and Facebook's rules against nudity ultimately gave way to a determination within the platform that it was "an iconic image of historical importance."[313] And when then-candidate Donald Trump's posts about his Muslim ban appeared to run afoul of Facebook's rules against hate speech, they stayed up after Mark Zuckerberg and senior members of Facebook's policy team concluded that they were "newsworthy, significant, or important to the public interest."[314] No doubt some newsworthiness judgments fall to policymakers who are lower on the Facebook totem pole, but the important point is that these decisions are made by humans—not algorithms—and often stem from company higher-ups.

---

311. Gertz v. Robert Welch, Inc., 418 U.S. 323, 346 (1974) (quoting Rosenbloom v. Metromedia, Inc., 403 U.S 29, 79 (1971) (Marshall, J., dissenting)).

312. *Id.*

313. Zillman, *supra* note 182.

314. Deepa Seetharaman, *Facebook Employees Pushed to Remove Trump's Posts as Hate Speech*, WALL ST. J. (Oct. 21, 2016, 7:43 PM), http://www.wsj.com/articles/facebook-employees-pushed-to-remove-trump-posts-as-hate-speech-1477075392 [https://perma.cc/63C2-LY4U].

Facebook's newsworthiness exception in some ways mirrors the courts' treatment of the same concept in privacy and IIED law. The interests served by both torts give way to a constitutional commitment to protecting robust discourse on matters of public concern, a determination that is often influenced by the judgment of the press.[315] But newsworthiness determinations at Facebook are made by an anonymous and somewhat arbitrary group of policy executives.[316] This infects the adjudication process with various problems and exposes a more systemic issue: the dearth of accountability that a private company like Facebook has to its users, despite impacting its users' speech rights on the platform.

The first problem created by ad hoc exceptions at platforms is that they might suffer from a lack of expertise or professional norms. At least in the courts, newsworthiness determinations are often shaped by deference to the traditional news media, where society has come to expect some level of skill, experience, and professionalism.[317] Even when tort claims do not involve a media defendant, judges themselves are usually proficient at reading and interpreting past decisions and deciding how they should apply in analogous contexts. It is at best unclear if we can expect the same from Facebook policymakers. Some may have expertise in content moderation, but others are there to represent the platform's business interests. Because users do not really know who is making these decisions, they are unable to judge what their skewed incentives might be.[318] But one thing is for sure: the decisions are currently made by Facebook *insiders*, and not by any neutral overseer less likely to prioritize the company's profits or prestige.

What little we do know about Facebook's decisionmaking process in this area is concerning. At least anecdotally, it seems that powerful and connected people are more likely to get favorable treatment. This materializes in two ways. For one, when prominent people complain about content takedowns, the platforms seem more likely to listen. When a Danish politician posted a photo of Copenhagen's iconic Little Mermaid statue on Facebook, the platform removed it for violating its nudity rules because the

315. *See, e.g.*, Gajda, *supra* note 236, at 1041.

316. Adler, *supra* note 129.

317. Even courts might have to become more skeptical of news deference in the digital age, in which everyone has a platform and professional norms of journalism are arguably deteriorating. *See* Gajda, *supra* note 236, at 1041–42; *cf.* Sarah C. Haan, *The Post-Truth First Amendment*, 94 IND. L.J. (forthcoming 2019).

318. *See* Sarah C. Haan, *Profits v. Principles*, KNIGHT FIRST AMEND. INST. (Oct. 30, 2018), https://knightcolumbia.org/content/profits-v-principles [https://perma.cc/6J89-U5ND] (observing that "[t]ensions exist between Facebook's business interests and its aspiration to create a prosocial expressive environment").

image featured "too much bare skin or sexual undertones."[319] After the politician complained publicly, Facebook backtracked and reinstated the post.[320] Similarly, after Facebook removed a post by a group of journalists in the Philippines criticizing President Rodrigo Duterte, the group was able to mobilize public pressure and persuade the platform to reinstate it.[321] Facebook may have provided everyone with a platform, but that does not mean that everyone is heard equally.[322]

The second way that platforms entrench power with the powerful is by allowing influential people to speak in ways that the general public cannot. Again, at least anecdotally, Facebook seems more likely to find the speech of public figures to be newsworthy, meaning that powerful users will often get a pass even when their content violates the platform's rules. Zuckerberg conceded as much when he justified the decision to treat Trump differently because his status as a "public figure" made the posts about the Muslim ban "newsworthy."[323] In the aftermath of that decision, Facebook announced that it would "begin allowing more items that people find newsworthy, significant, or important to the public interest—even if they might otherwise violate [its] standards."[324] This preferential treatment is worsened by what Jillian York has called the "tornadoes of celebrity," whereby different platforms reinforce each other in a way that creates new celebrities and then helps them remain powerful.[325]

Even though Facebook has now openly said that it makes exceptions for newsworthy content, it has given little insight into how it defines newsworthiness. Senior Facebook executive Peter Stern has said that the platform considers "safety of individuals on the one hand and voice on the other," but "voice" is a nebulous concept to say the least.[326] If policymakers

---

319. *Denmark: Facebook Blocks Little Mermaid Over 'Bare Skin,'* BBC (Jan. 4, 2016), https://www.bbc.com/news/blogs-news-from-elsewhere-35221329 [https://perma.cc/A69H-5XUH].

320. *Id.*

321. Aries Joseph Hegina, *Facebook Restores Journalists' Page with Anti-Duterte Post*, INQUIRER.NET (June 7, 2016, 11:15 AM), https://technology.inquirer.net/48270/facebook-restores-journalists-page-with-anti-duterte-post [https://perma.cc/92S4-FCCT].

322. *See* Press Release, Elec. Frontier Found., EFF, Human Rights Watch, and Over 70 Civil Society Groups Ask Mark Zuckerberg to Provide All Users with Mechanism to Appeal Content Censorship on Facebook (Nov. 13, 2018), https://www.eff.org/press/releases/eff-human-rights-watch-and-over-70-civil-society-groups-ask-mark-zuckerberg-provide [https://perma.cc/TR6L-TERT].

323. Seetharaman, *supra* note 314.

324. Kaplan & Osofsky, *supra* note 183.

325. Jillian York, Director of Int'l Freedom of Expression, Elec. Frontier Found., Speech at the re:publica 2018 Conference: The New Kingmakers: How Silicon Valley Created a New Culture of Celebrity (May 6, 2018), https://www.youtube.com/watch?v=Rl9JX5jmY0M [https://perma.cc/FGP6-CDJ7].

326. Telephone Interview with Peter Stern, *supra* note 184.

are making these calls on a case-by-case basis, without the benefit of a body of reasoned decisions from past cases, there is a risk that mere human whim can lead to erratic and arbitrary results. Of course, a nuanced and ambiguous standard like newsworthiness is always susceptible to inconsistent application, but Facebook has given us little confidence that it can reliably and fairly make these calls.

One cause for skepticism is the apparent American-centric conclusions that Facebook has drawn in this area. As we saw, the platform was quick to abandon its anti-gore rules to allow images of a Boston Marathon victim to remain online.[327] But when users made similar requests concerning videos depicting brutal violence committed by Mexican drug cartels, the result was quite different.[328] Lower-level Facebook employees deemed the Mexican videos newsworthy, but a high-ranking executive overruled them following intense backlash in the media.[329] The platform's rule seemed to shift depending on what side of the border the event occurred. Likewise, while Trump's Islamophobic posts about banning all Muslim immigration were saved due to their newsworthiness, Facebook removed posts by the son of Israeli prime minister Benjamin Netanyahu wishing that "all Muslims leave the land of Israel."[330]

This form of American exceptionalism is, in some sense, unsurprising. Facebook is an American company run largely by Americans, so a normative concept like newsworthiness has predictably been influenced by American norms. Yet what makes this trend at Facebook so interesting is that it seems contrary to how the leaders at the platform conceive of their mission. When Zuckerberg first raised the idea of creating a "Supreme Court" for content moderation, he extolled its virtue as a body that could "make the final judgment call on what should be acceptable speech in a community that *reflects the social norms and values of people all around the world.*"[331] This is an ambitious—and perhaps impossible—goal: the nature of norms and values is that they often develop differently within dissimilar communities. Facebook may wish that it could create one set of Community Standards that

327. Adler, *supra* note 129.

328. *Id.*

329. *Id.*

330. *Facebook Temporarily Bans Israeli PM's Son Over Posts*, BBC (Dec. 17, 2018), https://www.bbc.com/news/world-middle-east-46591270 [https://perma.cc/5FUE-KG2V]; *cf.* Jake Evans, *Fraser Anning's Public Facebook Page Removed for Reported Hate Speech*, ABC (Sept. 28, 2018, 12:53 AM), https://www.abc.net.au/news/2018-09-28/fraser-annings-facebook-page-taken-down/10317638 [https://perma.cc/MU6M-MUHJ] (discussing Facebook's banning of Australian politician from the platform after he violated rules prohibiting hate speech).

331. Klein, *supra* note 205 (emphasis added).

would satisfy everyone within its polity, but the reality is that these global platforms simply cannot craft policies that represent worldwide "norms and values" that do not exist—particularly surrounding such complex and contestable notions as the boundaries of free speech. Facebook will need to choose the values it wants to reflect in its content moderation, but the questions of how it will choose them and what those values might be remains unanswered.

In sum, several lessons emerge from a comparative analysis of courts' and platforms' treatment of public figures and newsworthiness. For starters, Facebook's use of Google News for its public-figure determinations underscores the dangers of reducing such judgments to mechanical calculations: when followed strictly, they can result in either taking down or keeping up too much speech. This approach, which necessarily jettisons the "voluntariness" caveat recognized by the courts, can lead to inaccurate and unjust results. Although Facebook and other platforms have proliferated the once-rare "involuntary public figure" imagined in *Gertz*, they—and the courts—have yet to reckon with this new reality. These issues reveal that the new speech ecosystem created by platforms like Facebook has indeed eroded the empirical postulate at the core of the public-figure doctrine—that is, if you are a public figure, you have greater access to channels of communication—now that everyone has a platform. Yet despite this apparent equalization of speech access, the result is often less than egalitarian. The powerful and connected often get better rules and treatment, not all speech receives equal attention or amplification, and the mechanisms used by Facebook to regulate speech—though increasingly transparent—remain largely opaque and unaccountable to users. Facebook's struggle to create principled exceptions for newsworthy content underscores how the company straddles the roles of the legislature, executive, judiciary, and press in controlling access to speech for both speakers and listeners.

CONCLUSION

As this Article reveals, platforms like Facebook have fundamentally altered the nature of the global speech ecosystem.[332] They have given

332. This Article has focused largely on Facebook for the purposes of comparing the private and public modes of adjudicating claims about harmful speech. But although Facebook is currently the most powerful and ubiquitous speech platform, it is of course just one of several players in this space. Content moderation is a key part of many internet platforms, from online marketplaces (for example, Etsy and Amazon) to gaming streaming (for example, Twitch and Steam), and from smaller niche communities (for example, Ravelry and Patreon) to global speech platforms (for example, Twitter and YouTube). The

everyone a platform to speak, dramatically reducing the barriers to entry in public debate. They have facilitated the amplification of speech, allowing users to reach broad audiences in places near and far. They have created both opportunities and dangers by enabling virality, accelerating and extending the influence of speech by previously obscure people. And they have developed a new system of governance, adjudicating the boundaries between protecting free speech and preventing harmful speech.

In the age dominated by the Old Governors, speech governance was essentially split between the legislature, executive, judiciary, and press.[333] At least in the United States, this governance system was divided between three branches of government, created and backed by values enshrined in the Constitution. Congress and state legislatures would make laws, those laws would be enforced by executive branches, and the legality of those laws would be measured against the Constitution by courts.[334] As a check against all of this official governance was the so-called "Fourth Estate" of the press, which enjoyed its own constitutional protection in its role as watchdog of the government and caretaker of an informed electorate.[335] The press's decisions of what to publish—what was "newsworthy"—were made by editorial boards and given some deference in courts.[336] Collectively, as Balkin has argued, this was the system of "old-school speech regulation."[337]

Today, in the age of the New Governors, we can see shadows of these various roles, but in a quite different construct. Much of the governance of online speech is done by private platforms that fulfill all of these roles—legislature, executive, judiciary, and press—at once. At Facebook, policy teams make rules on content, moderators enforce those rules in response to flagged content, and escalations teams review whether those enforcements

---

multitude of features offered by Facebook gives it a unique role in this new speech ecosystem and allows us to analogize to the various roles played by all of these actors, making its use as a case study particularly valuable. The lessons we can learn from Facebook's experience will be helpful to addressing parallel concerns raised by other online platforms. *See, e.g.*, *Twitter Will Hide Rule-Breaking Politicians' Tweets*, BBC (June 27, 2019), https://www.bbc.com/news/technology-48791094 [https://perma.cc/NPJ5-6UY2] (describing Twitter's new policy to hide politicians' "newsworthy" tweets that break the platform's rules but are left online "in the public interest").

333. *See* Balkin, *supra* note 8, at 2301, 2306–08.

334. *See* Marbury v. Madison, 5 U.S. (1 Cranch) 137 (1803).

335. As Edmund Burke once said, "there were Three Estates in Parliament; but, in the Reporters Gallery yonder, there sat a *Fourth Estate* more important far than they all." THOMAS CARLYLE, ON HEROES, HERO-WORSHIP, AND THE HEROIC IN HISTORY 202 (John Chester Adams ed., Houghton Mifflin & Co. 1907) (1842); *see also* Rachel Luberda, *The Fourth Branch of the Government: Evaluating the Media's Role in Overseeing the Independent Judiciary*, 22 NOTRE DAME J.L. ETHICS & PUB. POL'Y 507, 507–10 (2008).

336. Gajda, *supra* note 236, at 1041–42.

337. Balkin, *supra* note 8, at 2298.

were correct or if larger changes should be made in policy.[338] And while the media still plays an important role in raising awareness about problematic content-moderation rules or decisions,[339] platforms are also publishers of speech and act as editorial boards determining what types of content see the light of day. Platforms are both the governors, setting speech policies and adjudicating speech disputes, and the publishers, controlling access to speech on behalf of speakers and listeners. They are the *Sullivan* Court, and they are the *New York Times*.

Facebook seems to realize these parallels as well. Though it is just one of many speech platforms, it is leading the way in attempting to build accountability and oversight into its speech policies. By creating an independent "Oversight Board" to make policy and appeals determinations concerning the content users may post, Zuckerberg spoke directly to concerns raised by the consolidation of power at the platform.[340] Indeed, in his blog post making the announcement, he stated that creating the tribunal was an attempt to "prevent the concentration of too much decision-making within [its] teams" and to "provide assurance that these decisions are made in the best interests of [the] community and not for commercial reasons."[341] Concerns about Facebook's incentives have led some to worry that the body will be "more soundbite than substance,"[342] but there are promising signs that Zuckerberg will put his money where his mouth is—both literally and figuratively. In January 2019, the platform released a few more specifics about how the body might operate. In order to "render independent

338. Klonick, *supra* note 5, at 1630–58.

339. *See, e.g.*, Hegina, *supra* note 321.

340. Zuckerberg, *supra* note 123. The idea of such a tribunal did not spring from nowhere. Its genesis built on years of pressure from outside groups and Facebook's own attempts to create some form of appellate review for content-moderation decisions. In the 2015 Manila Principles, civil-society groups demanded that "[l]aws and content restriction policies and practices must respect due process" and provide "[t]ransparency and accountability." ELEC. FRONTIER FOUND., MANILA PRINCIPLES ON INTERMEDIARY LIABILITY 4–5 (2015), https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf [https://perma.cc/HNB9-F9EY]. David Kaye, the United Nations Special Rapporteur on Freedom of Expression, echoed these recommendations in his report to the Human Rights Council on free speech and the private sector in the digital age. David Kaye, *Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 89, U.N. Doc. A/HRC/32/38 (May 11, 2016) ("It is also critical that private entities ensure the greatest possible transparency in their policies, standards and actions that implicate the freedom of expression and other fundamental rights."). In a similar vein, the 2018 Santa Clara Principles declared that platforms "should provide a meaningful opportunity for timely appeal of any content removal or account suspension" and outlined several minimum standards that platforms should meet. THE SANTA CLARA PRINCIPLES ON TRANSPARENCY AND ACCOUNTABILITY IN CONTENT MODERATION (2018), https://santaclaraprinciples.org [https://perma.cc/K8ZN-7SNF].

341. Zuckerberg, *supra* note 123.

342. douek, *supra* note 204.

judgment" on the platform's speech policies, the body may "[r]everse Facebook's decisions when necessary."[343] According to the announcement, members of the board "will be obligated to the people who use Facebook—not Facebook the company."[344] Facebook even published a "Draft Charter" for the board, which included commitments that the board would "share its decisions transparently and give reasons for them" and that "Facebook will accept and implement the board's decisions."[345] Most significantly, the Charter also revealed that the board will base its decisions not only on Facebook's pre-existing Community Standards, but also on "a set of values, which will include concepts like voice, safety, equity, dignity, equality and privacy."[346] These values, which sound akin to constitutional provisions, will be included in a final charter "that will serve as the basis for board governance."[347]

Between January and June 2019, Facebook sought external advice on what the Oversight Board should look like.[348] Over those six months, it "heard from more than 650 people from 88 countries represented at 22 smaller global roundtables," gathered "feedback from more than 250 experts in one-on-one meetings," and established "an online 'public consultation' process, which encouraged users to both answer polls and submit essays on what they thought the board should look like."[349] As one might expect from such a global listening tour, the lengthy report released in late June summarizing the platform's findings raises more questions than answers.[350] But it does demonstrate Facebook's commitment to seeking out the opinions and thoughts of users and stakeholders worldwide and providing

---

343. Nick Clegg, *Charting a Course for an Oversight Board for Content Decisions*, FACEBOOK NEWSROOM (Jan. 28, 2019), https://newsroom.fb.com/news/2019/01/oversight-board [https://perma.cc/5ZA5-QG8L].

344. *Id.*

345. FACEBOOK, DRAFT CHARTER: AN OVERSIGHT BOARD FOR CONTENT DECISIONS 1 (2019), https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-1.pdf [https://perma.cc/EH7R-YZ7P].

346. *Id.* at 3.

347. *Id.* at 5.

348. Evelyn Douek & Kate Klonick, *Facebook Releases an Update on Its Oversight Board: Many Questions, Few Answers*, LAWFARE (June 27, 2019, 3:41 PM), https://www.lawfareblog.com/facebook-releases-update-its-oversight-board-many-questions-few-answers [https://perma.cc/7V2D-HFTV].

349. Kate Klonick & Evelyn Douek, *Facebook's Federalist Papers*, SLATE (June 27, 2019, 9:44 AM), https://slate.com/technology/2019/06/facebook-oversight-board-community-standards-federalist-papers.html [https://perma.cc/BBL9-5N34].

350. Douek & Klonick, *supra* note 348; *see also* Brent Harris, *Global Feedback and Input on the Facebook Oversight Board for Content Decisions*, FACEBOOK NEWSROOM (June 27, 2019), https://newsroom.fb.com/news/2019/06/global-feedback-on-oversight-board [https://perma.cc/C5PX-6866].

transparency in reporting those findings, which will hopefully influence the Oversight Board's final charter.[351]

Facebook is in the midst of its own kind of Constitutional Convention that could fundamentally alter its nature and the way it governs online speech.[352] Zuckerberg seems to have finally come to terms with his tremendous power, acknowledging that decisions about how to "balance safety and free expression . . . are too consequential for Facebook to make alone."[353] With the advent of an oversight body that aspires to bring "independent judgment to hard cases,"[354] the platform is on the cusp of creating a meaningful check on its own power that could have ripple effects throughout the industry and reshape public discourse on the internet. This may seem hyperbolic, but Facebook's 2.3 billion users leave a giant footprint on the character of online speech—and the nature of the platform's governance over them is among the most pressing issues concerning freedom of expression in the digital age.

Facebook's influence over online speech makes critical oversight all the more important. As the comparative analysis in this Article has shown, Facebook's approach to issues surrounding public figures and newsworthiness raises a host of problems. The platform's use of Google News to determine public-figure status is purely descriptive and lacks a normative backstop to consider concepts like voluntariness or community mores.[355] Yet if Facebook deviates from its algorithmic tools to make exceptions based on human judgment, it risks creating arbitrary and inconsistent results through an opaque process that is largely hidden from its users.[356] Facebook's exceptions for newsworthy content raise similar concerns.[357] Although the platform strives vaguely to balance the "voice" of its users against the "safety" of it users,[358] when it comes to difficult issues

351. For an insightful analysis of the potential benefits and limitations of Facebook's Oversight Board, see generally Evelyn Douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, 21 N.C. J.L. & TECH. 1 (2019).

352. *See generally* David Pozen, *Authoritarian Constitutionalism in Facebookland*, BALKINIZATION (Oct. 30, 2018), https://balkin.blogspot.com/2018/10/authoritarian-constitutionalism-in.html [https://perma.cc/AV6K-UYSC] (observing that Facebook's existing model is closer to "authoritarian constitutionalism" than a "common law system," in part because Facebook lacks "(i) formally independent dispute resolution bodies, paradigmatically courts, that issue (ii) precedential, (iii) written decisions").

353. Clegg, *supra* note 343.

354. *Id.*

355. *See supra* Sections III.B.1–.2.

356. *See supra* Section III.B.3.

357. *See supra* Section III.B.3.

358. Telephone Interview with Peter Stern, *supra* note 184.

surrounding "sympathetic" public figures and highly contextual newsworthiness determinations, the lack of transparent and granular articulations of the platform's decisions can lead to a host of problems.[359]

This Article has revealed the inner workings of Facebook's content moderation surrounding the crucial concepts of public figures and newsworthiness. The rules and processes that the platform has adopted have deep roots in First Amendment law, but they differ in critical respects. By comparing the old and new systems of speech governance, this Article has exposed flaws in both. But these flaws are not fatal—both judges and platform policymakers can change their doctrines to adapt to challenges posed by the new speech ecosystem brought about by companies like Facebook. The battle over how to protect free speech while regulating harmful speech must now be fought on two fronts: through tort law in courts and content moderation on platforms. While the Old Governors have long-established structures to adjudicate the public's claims, the New Governors are still building theirs.

---

359. *See supra* Section III.B.3.

✣ NOTES ✣