

NOTE
SOCIAL MEDIA DEFAMATION: A
NEW LEGAL FRONTIER AMID THE
INTERNET WILD WEST

HADLEY M. DREIBELBIS*

INTRODUCTION

Social media platforms have become the primary news source for many Americans. A 2019 Pew study found that over half of U.S. adults (55 percent) regularly rely on social media for their news.¹ Of these respondents, 52 percent received their news from Facebook, while YouTube was the second most popular platform at 28 percent, followed by Twitter and Instagram at 17 and 14 percent, respectively.² This evolution of social media as newsgathering often means sacrificing accuracy for expediency. Misinformation should be taken more seriously, because it can lead to devastating consequences. Posting false statements about people, even on social media platforms, can cause reputational ruin, financial damage, and lost economic opportunities. Social media “influencers,” government officials, and even ordinary users can cause lasting harm with the click of one button.

To rectify these injuries, courts have extended the concept of defamation beyond traditional media outlets like news publishers and broadcasters to include statements made online to hold reckless social media users responsible for their posts.³ But because defamation is

Copyright © 2021 Hadley M. Dreibelbis.

* J.D., Duke Law School, 2021; A.B., A.B., and A.B.J., University of Georgia, 2014. I am grateful to my faculty advisor, Professor Nicole Ligon, for her guidance and encouragement. My sincere appreciation to my Note Editor Catherine Tarantino and fellow editors at the *Duke Journal of Constitutional Law & Public Policy* for their thoughtful feedback. This Note is dedicated to my parents, Don and Anna Dreibelbis.

1. Elizabeth Grieco and Elisa Shearer, *Americans Are Wary of the Role Social Media Sites Play in Delivering News*, PEW RES. CTR. (Oct. 2, 2019), <https://www.journalism.org/2019/10/02/americans-are-wary-of-the-role-social-media-sites-play-in-delivering-the-news/>.

2. *Id.*

3. Nicole Pelletier, *The Emoji That Cost \$20,000: Triggering Liability for Defamation on Social Media*, 52 WASH. U. J. L. & POL'Y 227, 231 n.31 (2016).

state-based law, it is being applied in diverse ways, creating conflicting precedent. Some courts treat social media speech as per se hyperbole and exaggerated opinion, unable to be treated as an actionable statement of fact.⁴ On the other hand, other courts have rigidly applied traditional defamation principles, without understanding the unique context of social media as a casual, loose form of communication.⁵

This Note presents a navigable standard that recognizes that most users are unaware of the ramifications of their defamatory posts, while still holding them accountable for causing real harm to targeted individuals. Parts I & II discuss the development of defamation law and its First Amendment limitations. Part III looks at the courts' application of traditional defamation law to statements made on social media. It separates those courts that treat social media speech as opinion per se versus those who stringently apply a traditional defamation test, and it identifies some common traits of social media speech that help understand the court's reasoning. Finally, Part IV presents a limiting principle that should be applied to social media speech that balances the unique, freewheeling nature of the platforms with the need to hold users accountable for publishing false, damaging information.

I. THE FIRST AMENDMENT: PROTECTIONS AND EXCEPTIONS

The First Amendment protects the ability of Americans to speak freely and to associate with whomever they choose. This means that the government “has no power to restrict expression because of its message, its ideas, its subject matter, or its content.”⁶ Even highly offensive speech is protected, because the government “must remain neutral in the marketplace of ideas.”⁷ The United States adheres to the principle that “debate on public issues should be uninhibited, robust, and wide-open, and that it may well include vehement, caustic, and

4. See, e.g., *Feld v. Conway*, 16 F. Supp. 3d 1, 4 (D. Mass. 2014); *Glob. Telemedia Int'l, Inc. v. Doe 1*, 132 F. Supp. 2d 1261, 1270 (C.D. Cal. 2001); *Summit Bank v. Rogers*, 142 Cal. Rptr. 3d 40, 62 (Cal. Ct. App. 2012); *Matter of Konig v. WordPress.com*, 978 N.Y.S.2d 92, 94 (N.Y. App. Div. 2013); *Sandals Resorts Int'l Ltd. v. Google*, 925 N.Y.S.2d 407, 415–16 (N.Y. App. Div. 2011); *Jacobus v. Trump*, 51 N.Y.S.3d 330, 339 (N.Y. Sup. Ct. 2017).

5. See, e.g., *Unsworth v. Musk*, No. 2:18-cv-08048-SVW-JC, 2019 U.S. Dist. LEXIS 167626 at *13 (C.D. Cal. 2019); *D.C. v. R.R.*, 106 Cal. Rptr. 3d 399, 429 (Cal. Ct. App. 2010); *AvePoint, Inc. v. Power Tools, Inc.*, 981 F. Supp. 2d 496, 508–09 (W.D. Va. 2013); *SIO3, Inc. v. Bodybuilding.com, LLC*, No. CV 07-6311-EJL, 2008 U.S. Dist. LEXIS 129000 at *18 (D. Idaho 2008); *Bently Reserve LP v. Papaliolios*, 160 Cal. Rptr. 3d 423, 431 (Cal. Ct. App. 2013); and *Sanders v. Walsh*, 162 Cal. Rptr. 3d 188, 196 (Cal. Ct. App. 2013).

6. *Police Dep't of Chi. v. Mosley*, 408 U.S. 92, 95 (1972).

7. *Hustler Mag. v. Falwell*, 485 U.S. 46, 56 (1988) (quoting *FCC v. Pacifica Found.*, 438 U.S. 726, 745–46 (1978)).

sometimes unpleasantly sharp attacks on government and public officials.”⁸

Courts do recognize exceptions to this broad protection, however, in certain instances of defamatory speech or serious threats. Specific exceptions include “advocacy intended, and likely, to incite imminent lawless action; obscenity; defamation; speech integral to criminal conduct; so-called ‘fighting words’; child pornography; fraud; true threats; and speech presenting some grave and imminent threat the government has the power to prevent.”⁹ These exceptions have “no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.”¹⁰ When speech is considered outside the First Amendment, it may either be prohibited outright or significantly restrained. Defamation, defined as a false statement of fact, is considered one of the categories of speech which may be significantly restrained.

II. OVERVIEW OF DEFAMATION LAW

Defamation, a concept originating in 13th century English common law, aims to protect individuals from reputational harm stemming from false statements.¹¹ Historically speaking, the victims of defamation were often the Church or government—they had causes of action. Whereas today, defamation law primarily regulates behavior between private parties. While defamation law originally targeted seditious and blasphemous speech, it now concentrates on speech injurious to individual reputation.¹² These false statements are made either in writing (libel) or orally (slander) and are governed by state law.¹³ When defamation law was imported to America, the Founders did not recognize the tensions between this enshrinement of defamation law in the legal systems of American colonies and a burgeoning national constitutional framework protective of free expression.¹⁴ This seeming

8. *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 270 (1964).

9. *United States v. Alvarez*, 567 U.S. 709, 717 (2012).

10. *Chaplinsky v. New Hampshire*, 315 U.S. 568, 572 (1942).

11. See *Defamation*, BLACK’S LAW DICTIONARY (10th ed. 2014) (defining defamation as a “false written or oral statement that damages another’s reputation”).

12. See David L. Hudson Jr., *Libel and Slander*, THE FIRST AMEND. ENCYCLOPEDIA (May 14, 2020), <https://mtsu.edu/first-amendment/article/997/libel-and-slander> (stating that “the most important legal issue in a defamation case is determining the status of the plaintiff”).

13. *Defamation*, CORNELL L. SCH. LEGAL INFO. INST, <https://www.law.cornell.edu/wex/defamation> (last accessed Oct. 25, 2020).

14. See U.S. CONST. amend. I. (“Congress shall make no law respecting an establishment of

contradiction came to a head in the 1960s, when the Supreme Court for the first time applied the First Amendment to a defamation claim, thereby *constitutionalizing* defamation.¹⁵ After *New York Times v. Sullivan*, the First Amendment imposed heightened requirements of intent and greater protections for speech. This application created new tension between punishing false speech and protecting free expression.¹⁶ Later defamation cases further defined the distinctions between actual malice and negligence standards,¹⁷ public figures and private individuals,¹⁸ and what activities could be considered matters of public concern.¹⁹

To establish a *prima facie* case of defamation, there must be “(a) a false and defamatory statement concerning another; (b) an unprivileged publication to a third party; (c) fault amounting at least to negligence on the part of the publisher with respect to the act of publication; and (d) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.”²⁰ This test is particularly difficult to apply to speech made on a social media platform. To establish falsity, a plaintiff needs to distinguish fact from opinion, which can be difficult on a platform like Twitter where speech is generally hyperbolic, sometimes caustic, and prone to an “anything goes” attitude where users feel immune from consequences.²¹ Additionally, publication was once a barrier for many defaming parties due to the limited number of newspapers, media outlets, and broadcasters serving as traditional publishers.²² Internet and social media platforms, however, allow users to express themselves without the oversight of editorial gatekeepers. The Internet

religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.”).

15. *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 267 (1964).

16. *See id.* (holding that the First Amendment applies to defamation law).

17. *See Gertz v. Robert Welch*, 418 U.S. 323, 347 (1974) (“We hold that, so long as they do not impose liability without fault, the States may define for themselves the appropriate standard of liability for a publisher or broadcaster of defamatory falsehood injurious to a private individual.”).

18. *Id.*

19. *Time, Inc. v. Firestone*, 424 U.S. 448, 454 (1976).

20. RESTATEMENT (SECOND) OF TORTS § 558 (AM. LAW INST. 1977).

21. *Jacobus v. Trump*, 51 N.Y.S.3d 330, 339 (N.Y. Sup. Ct. 2017).

22. *See Elynn M. Angelotti, Twibel Law: What Defamation and Its Remedies Look Like in the Age of Twitter*, 13 J. HIGH TECH. L. 433, 452 (2013) (noting that the standardized publishing process limited the time institutions could publish).

democratized speech and thus expanded opportunities for exposure, but it also opened doors to defamation liability.

This Part explains how the evolution of defamation law led to the unique challenges facing social media speech today. First, it discusses the historical roots of defamation law in America. Next, it outlines the constitutionalization of libel law and the transition from criminal to civil liability. It then discusses the implications of applying the First Amendment to defamation law, including the development of an “actual malice” liability standard and distinctions between public figures, private individuals, and what constitutes matters of public concern. Finally, it highlights the specific challenges of applying defamation law in the social media context.

A. American Defamation Law’s Early Roots

The early origins of American defamation law can be traced directly to the English Star Chamber, which covered crimes of forgery, perjury, riots, maintenance, fraud, libel, and conspiracy.²³ Under the reign of King Henry VIII in the 16th century, the Star Chamber began focusing on prosecuting critics of the crown.²⁴ Even after the abolition of the Star Chamber in 1641, common law continued to cover prosecution of criminal libel.²⁵ Colonists then imported seditious libel to pre-independence America and prosecuted it vigorously under criminal law.²⁶ At that time, only proof that a defamatory statement had been spoken or written was all that was required for conviction; truth was not a defense.²⁷

The trial of John Peter Zenger in 1735 greatly influenced the development of the First Amendment of the U.S. Constitution. Zenger, the printer of the *New York Weekly Journal*, was prosecuted for seditious libel for criticism of New York Governor William Cosby.²⁸ At the time, seditious libel was defined as intentional publication of written criticism of public institutions, “without lawful excuse or

23. Martin Gruberg, *Star Chamber*, THE FIRST AMEND. ENCYCLOPEDIA (2009), <https://mtsu.edu/first-amendment/article/820/star-chamber>.

24. Eric P. Robinson, *Criminal Libel*, THE FIRST AMEND. ENCYCLOPEDIA (2009), <https://www.mtsu.edu/first-amendment/article/941/criminal-libel>.

25. *Id.*

26. *Id.*

27. Defamation in 18th century America meant the publication of statements “intended to criticize or provoke dissatisfaction with the government.” *Id.*

28. *Crown v. John Peter Zenger, 1735*, HIST. SOC’Y OF THE N.Y. CTS., <http://libdata.lib.ua.edu/login?url=https://history.nycourts.gov/case/crown-v-zenger/> (last accessed Mar. 11, 2021).

justification.”²⁹ The role of the jury was merely to decide whether Zenger was responsible for the allegedly libelous statement.³⁰ To the court’s surprise, Zenger’s counsel admitted Zenger’s responsibility, but he asked the jury to put aside the law and consider the truth of the statements published.³¹ The jury subsequently found Zenger “not guilty” of publishing seditious libel.³² Notwithstanding this advancement, states still continued to prosecute criminal libel under common law until the 1960s.

New York developed truth as a qualified defense to criminal libel in 1805, and a majority of states gradually followed suit.³³ Despite this legal evolution, however, defamation was still prosecuted as a strict liability offense, which meant that statements were presumed false and defendants bore the burden of proving truth.³⁴ In 1942, the Supreme Court affirmed that defamation was outside the bounds of protected First Amendment speech, stating that the prevention and punishment of libelous statements had “never been thought to raise any Constitutional problem.”³⁵ This strict separation of defamation and First Amendment screeched to a halt, however, in 1964.

B. The Constitutionalization of Defamatory Speech

In *Sullivan*,³⁶ the Supreme Court applied the First Amendment to a civil libel case for the first time and introduced the actual malice standard of intent to defamation law, significantly limiting the tort’s scope.³⁷ The case involved a newspaper ad defending Martin Luther King, Jr. and the civil rights movement.³⁸ The ad described police violence in Montgomery, Alabama, and a police commissioner sued the newspaper claiming the ad was defamatory.³⁹ The Supreme Court first addressed the issue of First Amendment protection by stating, “Like insurrection, contempt, advocacy of unlawful acts . . . and the various

29. *Id.*

30. *Id.*

31. *Id.*

32. *Id.*

33. Robinson, *supra* note 24.

34. See Alfred C. Yen, *It’s Not That Simple: An Unnecessary Elimination of Strict Liability and Presumed Damage In Libel Law*, 23 HARV. C.R.-C.L. L. REV. 593, 597 (1988) (“If the statement was defamatory, a defendant was held strictly liable at common law for actual harm proven at trial by the plaintiff.”).

35. 315 U.S. 568, 571–72 (1942).

36. 376 U.S. 254 (1964).

37. *Id.* at 269.

38. *Id.* at 257.

39. *Id.* at 256, 257.

other formulae for the repression of expression that have been challenged in this Court, libel can claim no talismanic immunity from constitutional limitations.”⁴⁰ The Court further asserted a commitment to “uninhibited, robust, and wide-open” debate on public issues, even if it may include “vehement, caustic, and sometimes unpleasantly sharp attacks on government and public officials.”⁴¹ “[E]rroneous statement is inevitable in free debate,” and free speech requires “breathing space” to survive.⁴²

To foster democratic debate, the Court required a higher level of intent when the allegedly defamatory speech involves a public official.⁴³ The Court emphasized the importance of treating government officials as “men of fortitude, able to thrive in a hardy climate,” because “[p]olitical conduct and views which some respectable people approve, and others condemn, are constantly imputed to [them].”⁴⁴ Thus, citizens and participants in the political process should be given wide latitude to discuss the characteristics and qualifications of those responsible for governing, and the Court ultimately held that speech about politicians and public officials is subject to the actual malice standard.⁴⁵ The actual malice standard, unlike the less rigorous negligence standard, requires the plaintiff to prove the statements were published with the knowledge that they were false or with a reckless disregard for the truth.⁴⁶

In the same year, the Supreme Court ruled in *Garrison v. Louisiana* that the actual malice standard would apply in criminal libel prosecutions as well.⁴⁷ The Court also recognized truth as an absolute defense.⁴⁸ This holding paved the way to the Court’s ruling in *Ashton v. Kentucky* in 1966, which effectively eliminated common law criminal libel and limited criminal prosecutions of libel to statutory law.⁴⁹ Presently, only fifteen states and U.S. territories have criminal libel statutes, and cases are rarely prosecuted.⁵⁰

40. *Id.* at 269.

41. *Id.* at 270.

42. *Id.* at 271–72.

43. *Id.* at 279–80.

44. *Id.* at 272 (quoting *Sweeney v. Patterson*, 128 F.2d 457, 458 (D.C. Cir. 1942)).

45. *Id.* at 281.

46. *Id.* at 280.

47. 379 U.S. 64, 79 (1964).

48. *Id.* at 73.

49. 384 U.S. 195, 201 (1966).

50. *Robinson*, *supra* note 24.

Soon after *Ashton*, the Court extended the actual malice standard to apply to public figures in addition to public officials in *Curtis Publishing Co. v. Butts*.⁵¹ Here, the plaintiff was a well-known University of Georgia athletic director who brought a libel action against the publisher of a magazine, *Saturday Evening Post*, for publishing an article accusing Butts of conspiring to “fix” a football game.⁵² Because Butts commanded “a substantial amount of independent public interest” at the time of publication, the Court considered him a public figure with sufficient access to the media and therefore had the ability to rebut the allegedly defamatory statements.⁵³ The Court ultimately held that a public figure must also prove actual malice on the part of publishers to recover damages for defamation.⁵⁴

C. Distinguishing Public Figures from Private Individuals

In the subsequent years, the Supreme Court further outlined the distinctions between public figures and private citizens and activities of public concern versus those of a private nature. These distinctions are critical, because they determine whether a plaintiff must prove actual malice or merely negligence to build a prima facie case of defamation.

The Court determined in *Gertz v. Robert Welch* that defamation cases involving private individuals require a lesser level of intent than actual malice.⁵⁵ In this case, the petitioner was an attorney who represented the victim’s family in a civil suit against a police officer convicted of his murder.⁵⁶ The respondent, a magazine publisher, published an article claiming the officer’s prosecution was part of a “Communist campaign against the police” and accused petitioner of being the “architect” of this conspiracy.⁵⁷

Rather than focusing on whether the trial was under public scrutiny, the Court instead defined the boundaries of when a plaintiff should be considered a public figure or official versus a private individual. Public figures “have thrust themselves to the forefront of particular public

51. 388 U.S. 130, 155 (1967).

52. *Id.* at 135.

53. *Id.* at 154–55.

54. *Id.* at 155.

55. *See* 418 U.S. 323, 347 (1974) (“We hold that, so long as they do not impose liability without fault, the States may define for themselves the appropriate standard of liability for a publisher or broadcaster of defamatory falsehood injurious to a private individual.”).

56. *Id.* at 325.

57. *Id.* at 325–26.

controversies in order to influence the resolution of the issues involved” and “invite attention and comment.”⁵⁸ The Court emphasized that truly involuntary public figures are “exceedingly rare.”⁵⁹ Because of this, publishers are entitled to presume that public officials and public figures have done so voluntarily and with knowledge of the increased risk of exposure that accompanies this attention.⁶⁰ This stands in contrast with private individuals, who have “relinquished no part of [their] interest in the protection of [their] own good name,” have less access to outlets to rebut defamatory statements, and are thus more vulnerable to injury.⁶¹ By contrast, public figures, due to their prominence in society, “usually enjoy significantly greater access to the channels of effective communication and hence have a more realistic opportunity to counteract false statements than private individuals normally enjoy.”⁶²

The Court also distinguished between two types of public figures: general-purpose and limited-purpose. General-purpose public figures have achieved “such pervasive fame or notoriety” to be a public figure in all contexts.⁶³ Alternatively, limited-purpose public figures have voluntarily injected themselves or been drawn into “a particular public controversy” and become a public figure for only a limited range of issues and duration.⁶⁴ When evaluating a limited-purpose public figure, courts must look to “the nature and extent of an individual’s participation in the particular controversy giving rise to the defamation.”⁶⁵ For example, celebrities like Brad Pitt, well-known musicians like Adele, and high-powered company executives like Bill Gates may be considered general-purpose public figures. On the other hand, limited-purpose public figures might include David Hogg, one of the Parkland survivors-turned-activists, or Casey Anthony, the controversial defendant acquitted for the murder of her child in 2011. In *Gertz*, the Court determined that the petitioner, acting as an attorney, was not a limited-purpose public figure, since he had not discussed the controversy with the press, had not engaged the public’s

58. *Id.* at 345.

59. *Id.*

60. *Id.*

61. *Id.*

62. *Id.* at 344.

63. *Id.* at 351.

64. *Id.*

65. *Id.* at 352.

attention, and did not “thrust himself into the vortex” of the public issue.⁶⁶ Consequently, he was a private individual.⁶⁷

The decision in *Time, Inc. v. Firestone* further narrowed the scope of the public figure distinction by holding that local notoriety alone does not qualify an individual as a public figure.⁶⁸ The defendant in the case was the publisher of *Time* magazine who wrote an article inaccurately claiming that plaintiff’s high-profile divorce was the result of extramarital affairs and extreme cruelty.⁶⁹ Defendant argued that the plaintiff, who had been married to the heir of the Firestone tire company fortune, was a public figure due to the public’s interest in the salacious details of an extremely wealthy couple’s marital difficulties.⁷⁰ In its brief, the defendant cited the District Court’s finding that “[d]ue to the social position of the participants and the sensational, colorful testimony at the trial, there was national news coverage” and that the plaintiff was a “central figure” in this controversy.⁷¹ The Court rejected this reasoning, saying the plaintiff “did not assume any role of especial prominence in the affairs of society, other than perhaps Palm Beach society.”⁷² Plaintiff was “compelled to go to court by the State” in order to obtain a divorce and did not freely choose to publicize details of her married life.⁷³ Otherwise, she was merely a private individual—Russell Firestone’s former wife and a “onetime Palm Beach schoolteacher.”⁷⁴ Not all controversies of interest to the public equate to “public controversies,” and local notoriety alone does not transform an individual into a public figure.

To be considered a public figure, the individual must also have *voluntarily* participated in public activities. Courts have considered, *inter alia*, the following activities to be “public activities”: advocating on national security issues, prominently opposing nuclear testing, posing as a Playboy Playmate, and professional belly dancing.⁷⁵ For

66. *Id.*

67. *Id.*

68. 424 U.S. 448, 454 (1976).

69. *Id.* at 451–52.

70. *Id.* at 454.

71. Brief for Petitioner at 32–34, *Time, Inc. v. Firestone*, 424 U.S. 448 (1976) (No. 74-944), 1975 WL 173764.

72. *Time, Inc. v. Firestone*, 424 U.S. 448, 453 (1976).

73. *Id.* at 454.

74. *Id.* at 452.

75. See e.g., *Pauling v. Globe-Democrat Publ’g Co.*, 362 F.2d 188, 196 (8th Cir. 1966); *Secord v. Cockburn*, 747 F.Supp. 779, 784 (D.D.C. 1990); *Vitale v. Nat’l Lampoon, Inc.*, 449 F. Supp. 442,

example, in a case concerning a researcher whose work was lambasted by a Member of Congress, the Court examined the researcher's activities to determine whether he was a limited-purpose public figure or private individual.⁷⁶ The Court ultimately held that he was a private individual, because he did not invite widespread public attention and comment by submitting federal grant applications or publishing his research.⁷⁷ Furthermore, his access to the media was not "regular and continuing"—he merely responded to the allegedly defamatory statement.⁷⁸ To the extent his research was of public concern, it was because of the Member's statement.⁷⁹ As the Court emphasized, "[T]hose charged with defamation cannot, by their own conduct, create their own defense by making the claimant a public figure."⁸⁰

These distinctions between general purpose public figures, limited-purpose public figures, and private individuals are critical for determining defamation liability, because they set the level of intent required for the defendant's allegedly defamatory statement. Actual malice, the standard for public figures, is much harder to overcome than negligence, the standard for private individuals.

D. Social Media's Unique Challenges to Defamation Law

The unique nature of social media platforms presents challenges to the application of traditional elements of defamation. First, a court must distinguish fact from opinion in circumstances where the two may appear virtually indistinguishable. It must also navigate the blurry line of who constitutes a public figure in the age of internet virality. Finally, a court will also need to reckon with the increased liability driven by the social media model of "users as content publishers."

To succeed on a defamation claim, plaintiffs must prove "falsity," which requires distinguishing fact from opinion.⁸¹ The Court separates this inquiry into two parts: a statement must be (1) "provable as false,"⁸²

445 (E.D. Pa. 1978); and *James v. Gannett Co.*, 353 N.E.2d 834, 840 (N.Y. 1976).

76. See *Hutchinson v. Proxmire*, 443 U.S. 111, 135–36 (1979) (narrowing the scope of limited-purpose public figure).

77. *Id.* at 135.

78. *Id.* at 136.

79. *Id.* at 135.

80. *Id.*

81. See *Gertz v. Robert Welch*, 418 U.S. 323, 339–40 (1974) ("Under the First Amendment there is no such thing as a false idea. However pernicious an opinion may seem, we depend for its correction not on the conscience of judges and juries but on the competition of other ideas.").

82. *Milkovich v. Lorain J. Co.*, 497 U.S. 1, 19 (1990).

and (2) “reasonably . . . interpreted as stating actual facts.”⁸³ The majority of state jurisdictions consider a totality of the circumstances under the second prong.⁸⁴ Courts may evaluate the medium through which the statement was published, the purpose of the communication, and the overall context.⁸⁵ Social media is often seen as more informal—a forum where users vent, get into heated debates, and crack jokes.⁸⁶ In an environment where social media users are viewed as more freewheeling or careless when speaking, the lines are blurred even further, and it becomes nearly impossible to distinguish opinions from factual statements during the falsity analysis. The unique nature of social media communication plays a significant role in evaluating whether, under the circumstances, the assertion at issue can be reasonably interpreted as a statement of fact.

As acknowledged by the line of defamation cases, public figures enjoy a unique platform that private individuals do not—they are able to easily reach the broader public due to their notoriety and have the ability to quickly respond to damaging statements through numerous media channels.⁸⁷ Additionally, public figures willingly assume the risks inherent in participating in public life.

However, with the advent of rapid-response social media platforms and the nature of Internet virality, private individuals now have access to the broader public, too. Average social media users may find themselves going “viral” over a humorous tweet or video and can parlay this attention into the development of a personal brand. For example, comedian Sarah Cooper became known on Twitter and TikTok in early 2020 for her lip-synching parodies of President Donald Trump; she then leveraged the exposure she gained from these videos to sign with a talent agency, land a Netflix comedy special, and even began developing a CBS television series.⁸⁸ Social media posts that gain

83. *Id.* at 20 (citing *Hustler Mag. v. Falwell*, 485 U.S. 46, 50 (1988)).

84. *See, e.g., McCabe v. Rattiner*, 814 F.2d 839, 842 (1st Cir. 1987); *Janklow v. Newsweek, Inc.*, 759 F.2d 644, 649 (8th Cir. 1985), *aff'd on reh'g*, 788 F.2d 1300 (8th Cir. 1986), *cert. denied*, 107 S. Ct. 272 (1986); *Scott v. News-Herald*, 25 Ohio St. 3d 243, 250 (Ohio 1986); *Info. Control Corp. v. Genesis One Computer Corp.*, 611 F.2d 781, 784 (9th Cir. 1980).

85. *Sandals Resorts Int'l Ltd. v. Google*, 925 N.Y.S.2d 407, 414 (N.Y. App. Div. 2011).

86. *Pelletier*, *supra* note 3, at 236.

87. The first remedy of any victim of defamation is self-help—using available opportunities to contradict the lie or correct the error and thereby to minimize its adverse impact on reputation. Public officials and public figures usually enjoy significantly greater access to the channels of effective communication and hence have a more realistic opportunity to counteract false statements than private individuals normally enjoy. *Gertz v. Robert Welch*, 418 U.S. 323, 344 (1974).

88. Megh Wright, *Sarah Cooper Has a TV Show in the Words Now Too*, *VULTURE* (Aug.

significant traction may also be republished by established media outlets, enshrining their perceived legitimacy. Finally, anyone with a social media account can rebut a damaging post with a “reply” or “comment,” and that response is permanently linked to the offending post. This democratization of social media access makes it all the more likely that any social media user could be perceived as at least a limited-purpose public figure, increasing the intent level required to prove defamation.

Finally, the publication process itself is impacted by the advent of social media. For traditional media outlets, the publishing process involves fact-checking, extensive copyediting, and even legal review.⁸⁹ Prior to the modern news era, these publishers and broadcasters also had the luxury of time—consistent distribution times, limited publication, and scheduled newscasts ensured that journalists had more time to develop—and verify—their stories.⁹⁰

Now, social media (particularly Twitter) has drastically changed the speed at which “breaking news” is disseminated. Audiences demand instantaneous reporting of information, which leaves almost no time to formally review to ensure accuracy. Social media users become newsgatherers and reporters themselves, filming public videos of incidents and sharing firsthand information that is later picked up by news outlets. Liabilities for false and defamatory statements thus spread to all publishers, not just those in the traditional media industry. Unlike traditional news editors, however, social media users are often unaware of the danger of posting content that harms another’s reputation.

III. COURT’S EVALUATION OF CONTEXT IN THE SOCIAL MEDIA AGE

Courts from California to New York have considered the role of defamation analysis in speech published on social media. Some courts treat social media speech as per se hyperbole, citing to factors like its casual, imprecise language, anonymity, and lack of editorial oversight.⁹¹

20, 2020), <https://www.vulture.com/2020/08/sarah-cooper-tv-show-cbs-comedy.html>.

89. Angelotti, *supra* note 22, at 451–52.

90. *Id.* at 452.

91. See, e.g., *Feld v. Conway*, 16 F. Supp. 3d 1, 4 (D. Mass. 2014); *Glob. Telemedia Int’l, Inc. v. Doe 1*, 132 F. Supp. 2d 1261, 1270 (C.D. Cal. 2001); *Summit Bank v. Rogers*, 142 Cal. Rptr. 3d 40, 62 (Cal. Ct. App. 2012); *Matter of Konig v. WordPress.com*, 978 N.Y.S.2d 92, 94 (N.Y. App. Div. 2013); *Sandals Resorts Int’l Ltd. v. Google*, 925 N.Y.S.2d 407, 415–16 (N.Y. App. Div. 2011); *Jacobus v. Trump*, 51 N.Y.S.3d 330, 339 (N.Y. Sup. Ct. 2017).

Other courts apply traditional defamation principles to social media speech, pointing to the recklessness with which users post statements and the specific knowledge claimed when asserting falsehoods.⁹² All courts recognize the unique nature of communication via social media platforms.

A. *One View: Social Media Speech Is Per Se Hyperbole*

Ever since social media exploded in the early 2000s with the development of Facebook, YouTube, MySpace, and Twitter,⁹³ courts have grappled with the repercussions of this loose, freewheeling style of communication on laws regulating false statements of fact. Internet users treat “status updates,” “stories,” “snaps,” and “tweets” as blips on the radar—temporary expressions of feelings and venting rather than statements with consequences. Recognizing this use, some courts have treated social media speech as per se hyperbole, meaning that exaggerated language and expressions of opinion cannot be pinned down as statements of fact.⁹⁴ Courts regularly cite the casual, venting nature of the speech, the possibility that the speech is published anonymously, and the lack of editorial standards to support their decision. Under this reasoning, social media speech cannot constitute defamation, and it is thus protected under the First Amendment. This treatment of social media speech has drawbacks, however, because it can lead to serious reputational harm without any viable means of legal recourse.

1. Casual, Imprecise Statements

New York courts are at the forefront of treating defamatory social media speech as per se hyperbole or opinion, and they often cite the informal nature of the speech. As Judge Ronnie Abrams articulated in a 2020 social media defamation decision, “If the Internet is akin to the

92. See, e.g., *Unsworth v. Musk*, No. 2:18-cv-08048-SVW-JC, 2019 U.S. Dist. LEXIS 167626 at *13 (C.D. Cal. 2019); *D.C. v. R.R.*, 106 Cal. Rptr. 3d 399, 429 (Cal. Ct. App. 2010); *AvePoint, Inc. v. Power Tools, Inc.*, 981 F. Supp. 2d 496, 508–09 (W.D. Va. 2013); *SIO3, Inc. v. Bodybuilding.com, LLC*, No. CV 07-6311-EJL, 2008 U.S. Dist. LEXIS 129000 at *18 (D. Idaho 2008); *Bently Reserve LP v. Papaliolios*, 160 Cal. Rptr. 3d 423, 431 (Cal. Ct. App. 2013); *Sanders v. Walsh*, 162 Cal. Rptr. 3d 188, 196 (Cal. Ct. App. 2013).

93. Irfan Ahmad, *The History of Social Media [Infographic]*, SOCIALMEDIATODAY (Apr. 27, 2018), <https://www.socialmediatoday.com/news/the-history-of-social-media-infographic-1/522285/>.

94. See, e.g., *Feld*, 16 F. Supp. 3d at 4; *Glob. Telemedia*, 132 F. Supp. 2d at 1270; *Rogers*, 142 Cal. Rptr. 3d at 62; *Konig*, 978 N.Y.S.2d at 94; *Sandals*, 925 N.Y.S.2d at 415–16; *Jacobus*, 51 N.Y.S.3d at 339.

Wild West . . . Twitter is, perhaps, the shooting gallery, where verbal gunslingers engage in prolonged hyperbolic crossfire.”⁹⁵ This crossfire blurs the line between fact and opinion, whether it happens on Twitter, online bulletin boards, or even over email.

In 2011, the New York Appellate Division considered a case between Sandals Resorts and Google, where Sandals sought to compel Google to disclose the identity of a Gmail account holder who sent emails criticizing Sandals Resorts.⁹⁶ The author of the email accused Sandals of prioritizing foreigners for senior managerial positions and relegating Jamaican nationals to menial jobs, all while receiving subsidies from the Jamaican government and being funded by Jamaican taxpayers.⁹⁷ The court considered the statement as a whole, the overall context of the email, and the tone and apparent purpose of the speech to determine whether or not it was a statement of fact.⁹⁸ It found that the email was an “exercise in rhetoric,” seeking investigation, and it had an angry and resentful tone.⁹⁹

The court then went further, stating that “[t]he culture of Internet communications, as distinct from that of print media such as newspapers and magazines, has been characterized as encouraging a ‘freewheeling, anything-goes writing style.’”¹⁰⁰ It also emphasized that Internet forums contain a “wide range of casual, emotive, and imprecise speech” which are riddled with “grammatical and spelling errors, the use of slang, and, in many instances, an overall lack of coherence.”¹⁰¹ The court ultimately held that while portions of the email could arguably be considered statements of fact, these portions “constitute facts supporting the writer’s opinion,” which rendered the entire email “pure opinion” not actionable under defamation law.¹⁰²

Similarly, the court in *Global Telemedia International, Inc. v. Doe 1* held that pseudonymous messages on an Internet forum did not constitute defamation since they were considered per se hyperbole.¹⁰³

95. Ganske v. Mensch, No. 19-CV-6943, 2020 WL 4890423 at *1 (S.D.N.Y. Aug. 20, 2020).

96. *Sandals*, 925 N.Y.S.2d at 409.

97. *Id.*

98. *Id.* at 414.

99. *Id.* at 415.

100. *Id.* (citing Eirik Cheverud, Comment, Cohen v. Google, Inc., 55 N.Y.L. SCH. L. REV. 333, 335 (2010)).

101. *Id.* at 415–16 (citing Jennifer O’Brien, Note, *Putting a Face to a (Screen) Name: The First Amendment Implications of Compelling ISPs to Reveal the Identities of Anonymous Internet Speakers in Online Defamation Cases*, 70 FORDHAM L. REV. 2745, 2774–75 (2002)).

102. *Id.* at 416.

103. 132 F. Supp. 2d 1261, 1267 (C.D. Cal. 2001).

Using pseudonyms, defendants posted numerous messages on an Internet bulletin board referencing the plaintiff company in a “less-than-flattering” manner.¹⁰⁴ The court held that the statements were nonactionable opinion, in part because they were “full of hyperbole, invective, short-hand phrases and language not generally found in fact-based documents, such as corporate press releases or SEC filings.”¹⁰⁵ For example, statements about the company’s investors also contained phrases like “get off my back cowboy,” “grow up kids before you fall off your perch,” and “gotta love this companies [sic] potential.”¹⁰⁶ The postings “lack[ed] the formality and polish typically found in documents in which a reader would expect to find facts.”¹⁰⁷

Another case involving Internet message boards, *Summit Bank v. Rogers*, featured posts that the court considered to be “free-flowing diatribes (or ‘rants’) in which [defendant] does not use proper spelling or grammar, and which strongly suggest that these colloquial epithets are his own unsophisticated, florid opinions.”¹⁰⁸ The court noted that the posts appeared in Craigslist’s “Rants and Raves” section, which should predispose the reader “to view them with a certain amount of skepticism, and with an understanding that they will likely present one-sided viewpoints rather than assertions of provable facts.”¹⁰⁹

Courts also point to the argumentative and heated nature of social media debate as another reason to believe statements are per se opinion and not factual claims. In *Matter of Konig v. Wordpress.com*,¹¹⁰ the court stated that given the context of speech made “on an Internet blog during a sharply contested election,” a respondent’s reference to “downright criminal actions” would reasonably be viewed as a statement of opinion rather than a factual criminal accusation.¹¹¹ Additionally, the court in *Feld v. Conway* noted that the allegedly defamatory tweet was “made as part of a heated Internet debate” before concluding that a reasonable person would not understand the tweet to contain actual facts.¹¹²

104. *Id.* at 1264.

105. *Id.* at 1267.

106. *Id.*

107. *Id.*

108. 142 Cal. Rptr. 3d 40, 62 (Cal. Ct. App. 2012).

109. *Id.* at 60.

110. 978 N.Y.S.2d 92 (N.Y. App. Div. 2013).

111. *Id.* at 94.

112. 16 F. Supp. 3d 1, 4 (D. Mass. 2014).

The strongest pronouncement that social media speech is per se opinion arrived with the 2017 case, *Jacobus v. Trump*.¹¹³ The plaintiff, public relations consultant Cheri Jacobus, was approached by Donald Trump’s presidential campaign to discuss her potential employment as the campaign’s communications director.¹¹⁴ Initially, Jacobus showed interest in the position. After an argument erupted during one of her meetings with campaign officials, however, she withdrew her candidacy.¹¹⁵ Months later, Jacobus criticized then-candidate Trump on television.¹¹⁶ Consequently, Trump published a series of tweets on Twitter that claimed Jacobus “begged” his campaign for a job, and that she “went hostile” after being turned down twice for the position.¹¹⁷ As a result, Jacobus faced vile personal attacks by Trump’s numerous Twitter followers.¹¹⁸ In response, she filed a defamation suit against Trump that claimed his tweets cost her professional opportunities and injured her reputation as a political commentator.

The New York Superior Court first looked to the context in which the statements were made, including “the words themselves and their purpose, the circumstances surrounding their use, and the manner, tone, and style with which they are used.”¹¹⁹ It further explained that statements made “during the course of a heated public debate” are not actionable—even if appearing to be statements of fact—because an audience would expect use of hyperbolic language and rhetoric.¹²⁰ Finally, the court confirmed that it was following a consistent precedent of protecting statements made in online forums as statements of opinion rather than fact.¹²¹ Although it recognized the significant harassment Jacobus received from Trump’s followers, her reputational injury, and her loss of professional and economic opportunities, the

113. 51 N.Y.S.3d 330 (N.Y. Sup. Ct. 2017).

114. *Id.* at 333.

115. *Id.* at 334.

116. *Id.*

117. *Id.* Former President Trump’s Twitter account has since been permanently suspended due to “risk of further incitement of violence” after the insurrectionist storming of Capitol Hill on January 6, 2021. See *Permanent Suspension of @realDonaldTrump*, TWITTER INC. (Jan. 8, 2021), https://blog.twitter.com/en_us/topics/company/2020/suspension.html.

118. See *id.* at 334–35 (stating that Jacobus was attacked through demeaning tweets, including “an image of plaintiff with a grossly disfigured face,” and a “depiction of her in a gas chamber with Trump standing nearby ready to push a button marked ‘Gas’”).

119. *Jacobus*, 51 N.Y.S.3d at 336 (quoting *Steinhilber v. Alphonse*, 68 N.Y.2d 283, 291–92 (N.Y. 1986)).

120. *Id.* 336–37.

121. *Id.* at 339.

court held Trump’s statements to be nonactionable opinions protected by the First Amendment.¹²²

In August 2020, *Ganske v. Mensch* involved conspiracy theorist Louise Mensch, who tweeted allegedly defamatory statements about former Associated Press (AP) journalist Charles Ganske accusing him of xenophobia and of “clearly personally spread[ing] Russian bots.”¹²³ Soon after this Twitter exchange, the AP fired Ganske.¹²⁴ Despite the professional fallout from this exchange, the *Ganske* court granted the defendant’s motion to dismiss, relying on *Jacobus* for precedent and holding that a tweet calling someone “xenophobic” was nonactionable opinion.¹²⁵ The court determined that the tweet’s language amounted to fiery rhetoric and a personal reaction to the plaintiff’s own words, rather than a conveyance of objective facts.¹²⁶ Defendant Mensch also included a hyperlink in her tweet to the alleged bot activity.¹²⁷ The court treated Mensch’s linking to a data set as laying a factual basis for her assertion and held it to be nonactionable.¹²⁸ Finally, Mensch tweeted that Ganske was sent into a “frenzy,” which the court said has no precise meaning.¹²⁹

2. Anonymous Speech

Courts also emphasize the often-anonymous nature of social media speech in categorizing it as per se opinion. Generally, plaintiffs who bring a defamation lawsuit must accurately name a defendant to state their claims.¹³⁰ But social media platforms protect user identities (assuming users don’t themselves disclose personally identifying information) and only disclose them when dealing with circumstances like compliance with a court order.¹³¹ Therefore, unless plaintiffs convince a court to force platforms like Facebook or Twitter to disclose a user’s identity, they may not get very far with their defamation claims in the first place.

122. *Id.* at 343.

123. *Ganske v. Mensch*, No. 19-CV-6943, 2020 WL 4890423 at *6 (S.D.N.Y. Aug. 20, 2020).

124. *Id.* at *3.

125. *Id.* at *7.

126. *Id.*

127. *Id.* at *6.

128. *Id.*

129. *Id.* at *7.

130. Cory Batza, Comment, *Trending Now: The Role of Defamation Law in Remediating Harm from Social Media Backlash*, 44 PEPP. L. REV. 429, 453 (2017).

131. *Id.*

When a social media user only identifies by an anonymous username, not only will plaintiffs have a difficult time naming a defendant, but the courts may also be less likely to treat the statements made by that social media user as actionable statements of fact. For instance, in *Krinsky v. Doe 6*, pseudonymous accounts posted scathing critiques of a company on an Internet message board.¹³² Ultimately finding that these accounts merely posted nonactionable opinions and rhetoric, the court discussed “the relative anonymity afforded by the Internet forum [that] promotes a looser, more relaxed communication style.”¹³³ Users may be more provocative, combative, and free to engage in criticism without fear of retribution.¹³⁴ Thus, Internet discussions begin to look more like “vehicle[s] for emotional catharsis” than sources of news and information.¹³⁵ The court added that the First Amendment has long protected the right to publish anonymously, in order to encourage ideas to enter the marketplace freely.¹³⁶

Both *Sandals* and *Global Telemedia* addressed the relationship between anonymity and alleged defamation. The *Sandals* court railed against disclosure of anonymous speakers’ identities, expressing concern that corporations and wealthy plaintiffs could use subpoenas to silence online critics through public identification.¹³⁷ In *Global Telemedia*, the court stated that the proliferation of anonymous accounts created both a “general cacophony” and an “ongoing, free-wheeling, and highly animated exchange” in the Internet chat room.¹³⁸ The court in *Summit Bank* emphasized that “the very fact that most of the posters remain anonymous, or pseudonymous, is a cue to discount their statements accordingly.”¹³⁹ Not only did the court’s statement refute the important First Amendment protection of anonymity as a

132. 72 Cal. Rptr. 3d 231, 234 (Cal. Ct. App. 2008).

133. *Id.* at 237–38.

134. *Id.* at 238.

135. *Id.*

136. See *McIntyre v. Ohio Elections Comm’n*, 514 U.S. 334, 342 (1995) (“[A]n author’s decision to remain anonymous, like other decisions concerning omissions or additions to the content of a publication, is an aspect of the freedom of speech protected by the First Amendment.”). Indeed, authors publishing under pseudonyms have a long history in America. Perhaps the most notable example from American history is the Federalist Papers, written by John Jay, Alexander Hamilton, and James Madison, under the collective pseudonym “Publius.” Later, Hamilton and James Madison debated Washington’s Neutrality Proclamation under the pseudonyms “Pacifcus” and “Helvidius,” respectively. See *Pacifcus/Helvidius Letters*, George Washington’s Mount Vernon, <https://www.mountvernon.org/library/digitalhistory/digital-encyclopedia/article/pacifcus-helvidius-letters/> (last accessed Nov. 8, 2020).

137. *Sandals Resorts Int’l Ltd. v. Google*, 925 N.Y.S.2d 407, 416 (N.Y. App. Div. 2011).

138. *Glob. Telemedia Int’l, Inc. v. Doe 1*, 132 F. Supp. 2d 1261, 1267 (C.D. Cal. 2001).

139. *Summit Bank v. Rogers*, 142 Cal. Rptr. 3d 40, 60 (Cal. Ct. App. 2012).

means of advancing democratic debate; it also helped future defendants easily rebuff defamation claims.

3. No Editorial Gatekeeping

The third element courts often consider when determining social media speech to be per se opinion is the lack of editorial gatekeepers on Internet platforms.

As the *Sandals* court noted, “The low barrier to speaking online allows anyone with an Internet connection to publish his thoughts, free from the editorial constraints that serve as gatekeepers for most traditional media of disseminating information.”¹⁴⁰ These publishing guardrails established mechanisms for fact-checking and accuracy determinations that are often absent in rapid-fire social media publishing. Editorial gatekeepers also limit the number of publishers with access to media. This gatekeeping is now primarily relegated to traditional media platforms, as “anyone who has access to the Internet has access to . . . chat-rooms.”¹⁴¹ Social media platforms do not place constraints on publishing beyond their “community standards,” nor do they require any “special expertise, knowledge, or status” to post, tweet, or reply.¹⁴²

Another case, *John Doe No. 1 v. Cahill*, involved pseudonymous statements posted on a community blog.¹⁴³ Here, the court described the Internet as a “unique democratizing medi[a]” which allowed increasingly diverse people to engage in public debate.¹⁴⁴ Speakers can “bypass mainstream media to speak directly” to an audience, and anyone can “become a town crier with a voice that resonates farther than it could from any soapbox.”¹⁴⁵ The court also found that this universal access allows individuals targeted by speech to directly correct the record without barriers.¹⁴⁶ This ability counterbalances the protections that might have been potentially lost with erasure of the editorial process.

140. *Sandals*, 925 N.Y.S.2d at 415 (citing O’Brien, *supra* note 101, at 2774–75).

141. *Glob. Telemedia*, 132 F. Supp. 2d at 1264.

142. *Id.*

143. 884 A.2d 451 (Del. 2005).

144. *Id.* at 455.

145. *Id.*

146. *Id.*

B. Opposing View: Social Media Speech Is Subject To Traditional Defamation Analysis

Many other courts—particularly those in California—treat social media speech no differently than traditional media statements, applying the same defamation analysis to both. As more users turn to social media for breaking news, thoughtful analysis, and even engagement with traditional media outlets, social media users must ensure statement accuracy to prevent spreading damaging, false information. The court in *SIO3, Inc. v. Bodybuilding.com* emphasized this view, stating that “[i]f the Internet is to be considered a ‘marketplace of ideas’ and lent a commensurate air of trust and reliability, ‘it can never achieve its potential as such unless it is subject to the civilizing influence of the law like all other social discourses.’”¹⁴⁷ These courts that apply traditional defamation analysis primarily focus on three characteristics of social media speech in determining liability: (1) a serious tone; (2) claims of specialized knowledge, research, or news; and (3) specificity.

1. Serious Nature of Remarks

Courts are more willing to hold a social media user liable for defamation for statements made with a serious tone, using proper grammar and spelling.

A case covered extensively in headlines, *Unsworth v. Musk*, considered tweets that Elon Musk published about a cave diver involved in the 2011 rescue of a child soccer team in Thailand.¹⁴⁸ When the diver described Musk’s attempt to build a mini-submarine to rescue the boys as a “PR stunt,” Musk fired back on Twitter, accusing the rescuer of being “sus” (shorthand for suspicious), and a “pedo guy” (a common abbreviation for pedophile).¹⁴⁹ He then sent an email to a BuzzFeed reporter doubling down on these claims.¹⁵⁰ Although Musk argued that his tweets should be viewed with a “heavy presumption” of opinion due to Twitter’s unfiltered nature, the court rebutted this

147. *SIO3, Inc. v. Bodybuilding.com, LLC*, No. CV 07-6311-EJL, 2008 U.S. Dist. LEXIS 129000 at *18 (D. Idaho 2008) (citing *Varian Med. Sys., Inc. v. Delfino*, 6 Cal. Rptr. 3d 325, 337 (Cal. Ct. App. 2003) (rev’d on alternate grounds)).

148. *Unsworth v. Musk*, No. 2:18-cv-08048-SVW-JC, 2019 U.S. Dist. LEXIS 167626 at *2 (C.D. Cal. 2019).

149. *Id.* at *3–4.

150. *See id.* at *7 (disclosing the specific text used in Musk’s email).

argument by noting that Twitter is now an important source of facts in addition to opinions.¹⁵¹

Musk then argued that his tweets contained “imaginative and non-literal” insults, and that his joking manner was obvious from his use of both shorthand and colloquialisms.¹⁵² The court acknowledged that some prior cases found loose, hyperbolic language to be nonactionable opinion.¹⁵³ Nonetheless, it said that Musk’s tweets used “generally proper grammar with punctuation and had very few misspellings.”¹⁵⁴ Furthermore, Twitter’s 280-character limit often requires users to employ abbreviations and shorthand phrases. Therefore, the use of shorthand phrases should not automatically discount a tweet as opinion rather than fact.¹⁵⁵ Ultimately, the court determined that Musk’s tweets contained facts susceptible to verification, because the “[p]laintiff either is a pedophile or he is not and, if he were, evidence could prove it.”¹⁵⁶

D.C. v. R.R. also examined the seriousness and tone of social media speech in the defamation context.¹⁵⁷ In this case, the plaintiff—a student with an acting and singing career—maintained a website with an open “guestbook” for members of the public to post comments.¹⁵⁸ After several students at his school posted derogatory comments and threats on the website, D.C. sued them over libelous statements about his sexual orientation.¹⁵⁹ Although the court primarily discussed these statements under its analysis of a “true threat,” its language about online speech is enlightening. According to the court, these statements were not merely “a few words shouted during a brawl”; rather, they constituted a “series of grammatically correct sentences composed at a computer keyboard.”¹⁶⁰ The author had to deliberate before pressing the “send” button, giving him time to retract the words.¹⁶¹ Thus, the court deduced the statement’s tone from the students’ grammatical choices.

151. *Id.* at *11.

152. *Id.* at *15–16.

153. *Id.* at *15.

154. *Id.* at *16.

155. *Id.*

156. *Id.* at *20.

157. 106 Cal. Rptr. 3d 399, 420 (Cal. Ct. App. 2010).

158. *Id.* at 405–06.

159. *Id.* at 406.

160. *Id.* at 420.

161. *Id.*

2. Claims of Specialized Knowledge

Courts are also more likely to evaluate social media speech under traditional defamation analysis when the poster claims to be “unbiased,” have “specialized knowledge,” or calls his posts “Research Reports,” bulletins, or alerts.¹⁶² For instance, in *AvePoint, Inc. v. Power Tools, Inc.*, employees of a rival company engaged in a series of false and deceptive claims about its competitor on Twitter and in direct customer communications.¹⁶³ The Twitter statements insinuated that AvePoint was a Chinese company, using the hashtag “MADEINCHINA” and referring to the company as the “Red Dragon.”¹⁶⁴ AvePoint relied on business with the federal government, who must give preferential treatment to domestic production.¹⁶⁵ The rival company also claimed its product was “Microsoft recommended” over AvePoint software, which could have prejudiced AvePoint’s business.¹⁶⁶ When employees of a competitor company post statements about their rival, they are speaking with the knowledge derived from their specialized industry—in this case, software. Therefore, the court found that potential customers were more likely to view these statements as serious assertions of fact.¹⁶⁷

SI03 v. Bodybuilding.com also touched on Internet speech that spoke to company practices.¹⁶⁸ Acknowledging that it is often difficult to determine the line between actionable statements of fact and protected opinion, the court said, “False statements about a company and/or its products literally killing people go beyond the general distasteful nature that is, at times, dominant within an electronic forum like Internet chat rooms.”¹⁶⁹ Even though the statements were riddled with colloquialisms and grammatical mistakes, they still had potential to mislead customers and harm the company’s reputation, making them actionable under defamation law.

162. See, e.g., *Overstock.com, Inc. v. Gradient Analytics, Inc.*, 61 Cal. Rptr. 3d 29, 42 (Cal. Ct. App. 2007) (holding defendant liable for defamation due to its claimed “specialized knowledge in the areas of financial accounting and issues of earnings” and publications titled “Research Reports”).

163. 981 F. Supp. 2d 496, 503 (W.D. Va. 2013).

164. *Id.* at 507.

165. *Id.*

166. *Id.* at 508.

167. *Id.* at 509.

168. *SI03, Inc. v. Bodybuilding.com, LLC*, No. CV 07-6311-EJL, 2008 U.S. Dist. LEXIS 129000 at *1 (D. Idaho 2008).

169. *Id.* at *32.

3. Specificity Means Provability

Statements made with specificity are also more likely to be considered statements of fact. Notably, many of the cases where Internet speech is considered specific and actionable involve business reviews on Yelp and other online message boards, where customers may be misled or real economic consequences could be felt.

The *Unsworth* court not only discussed the serious nature of Musk's statements, but also their specificities.¹⁷⁰ In Musk's emails to BuzzFeed, he made additional specific allegations of pedophilia against the plaintiff.¹⁷¹ These included, in part, that "[h]e's an old, single white guy from England who[] . . . mov[ed] to Chiang Rai for a child bride who was about 12 years old at the time."¹⁷² The court concluded that the email was written in "clear, plain, and non-figurative language."¹⁷³ It also found that Musk alleged "highly detailed facts that, due to their specific nature, could be readily verified."¹⁷⁴ Furthermore, the court determined that his tweet that offered a "signed dollar" to anyone who could confirm the cave rescuer's alleged pedophilia did not communicate "uncertainty"; instead, it doubled down on Musk's willingness to "bet on [the statement's] veracity."¹⁷⁵

In *Bently Reserve LP v. Papaliolios*, the court examined a defendant's Yelp review of an apartment building.¹⁷⁶ The defendant stated in his review that the new owners of the building likely "contributed to the death[s] of three tenants. . . and the departure[s] of eight more."¹⁷⁷ He also claimed they sought evictions of six long-term tenants without cause, and that he had "personally witnessed" the owners's "abhorrent behavior."¹⁷⁸ The court held these statements to be "purported facts," even though the posts also contained hyperbolic language.¹⁷⁹ It distinguished the precedents of *Krinsky* and *Summit Bank* as cases consisting of "true rants and raves," while arguing that the cases did not "preclude the courts from taking serious Internet speech seriously."¹⁸⁰ Furthermore, although generalized comments may

170. No. 2:18-cv-08048-SVW-JC, 2019 U.S. Dist. LEXIS 167626 at *3 (C.D. Cal. 2019).

171. *See id.* at *6–7.

172. *Id.*

173. *Id.* at *17.

174. *Id.* at *20.

175. *Id.* at *18–19.

176. 160 Cal. Rptr. 3d 423, 426 (Cal. Ct. App. 2013).

177. *Id.*

178. *Id.*

179. *Id.* at 430.

180. *Id.* at 433.

indicate expressions of opinion, “specifics, if given, may signal the opposite and render an Internet posting actionable.”¹⁸¹

That same year, a California state court considered a defamation suit involving another Yelp review. In *Sanders v. Walsh*, the court analyzed anonymous statements made about a business owner.¹⁸² The defendants posted a series of statements on “RipoffReport.com,” claiming that plaintiff used an “unauthorized” check to purchase a wig from the store and forged a FedEx letter to show an attempted return.¹⁸³ Here, the court held that the statements were not “vague implications” of fact but were instead “specific factual claims.”¹⁸⁴ It rested this conclusion on two bases: First, the statements about the plaintiff’s city contracts were prefaced by “Fact:,” and alleged historical claims of perjury and fraud.¹⁸⁵ Second, these statements—such as one alleging that the plaintiff gave “all the construction business in Anaheim for a[n] under the table bribe”—could be proven true or false.¹⁸⁶ Therefore, the court found that the specificity of the statements made them provable and thus actionable under defamation law.

IV. WHO HAS IT RIGHT?

Social media has become an increasingly globalized communications tool and an important platform for delivering breaking news. Courts should hold social media users responsible for the statements they post because these statements could be viewed as serious, newsworthy, and consequential due to the prominence of these platforms. Defamation law should not apply a rigid test, however, that fails to consider the instantaneous, freewheeling nature of social media speech. Courts should instead consider a more balanced approach. First, they should consider the context of the speech in their falsity analyses, including whether the statements made with exaggerated language, casual conversation, and heated rhetoric. Consequently, social media statements made with specificity and serious tone would be treated more like actionable statements of fact rather than per se hyperbole. Any resulting potential for increased liability from the falsity analysis could be counterbalanced by heightening the requisite

181. *Id.*

182. 162 Cal. Rptr. 3d 188, 192 (Cal. Ct. App. 2013).

183. *Id.*

184. *Id.* at 196.

185. *Id.*

186. *Id.*

level of intent for alleged defamers to actual malice—particularly when plaintiffs are “influencers” with large social media followings. Courts should treat all online platforms consistently when conducting this analysis, recognizing that Twitter and blogs are no longer mere casual conversation platforms.

A. *Finding A More Balanced Approach*

Courts must apply the traditional elements of defamation to the facts of a social media case. By treating social media speech as per se opinion or hyperbole, courts are undervaluing the impact that this speech has on individuals’s reputations and economic opportunities. There are currently 330 million monthly active users on Twitter,¹⁸⁷ 2.45 billion on Facebook,¹⁸⁸ and over one billion on Instagram.¹⁸⁹ Users of these sites rely on accurate information to make purchasing decisions, stay informed about breaking news, and share reliable data with their personal networks of family and friends. These sites are no longer mere outlets solely for blowing off steam or sharing personal opinions; they are increasingly e-commerce and news platforms too. Speech made on these social media platforms must be taken seriously.

Courts may, however, look to the context in which a statement is made on social media to determine whether there is an actionable false statement of fact. Opinions, “hot takes,” and casual conversations are just as critical as statements of fact, breaking news, and detailed serious language to social media communications. In order to determine whether a reasonable user would believe the individual was making an actionable statement of fact, any application of the defamation test must consider the speech’s specific facts, tone, and language. Social media platforms should not be independently evaluated as per se “opinion” sites; instead, courts should conduct a context-specific analysis, looking to the broader conversation, debate, or thread.

For example, say one Twitter user gets into an argument with another over a political disagreement and posts, “@TwitterUser2 roots for death and destruction! He wants to see this country ruined.” These statements, which may be hurtful to @TwitterUser2, are exaggerations

187. Ying Lin, *10 Twitter Statistics Every Marketer Should Know in 2020 [Infographic]*, OBERLO (May 30, 2020), <https://www.oberlo.com/blog/twitter-statistics>.

188. Katie Sehl, *27 Top Facebook Demographics that Matter to Social Media Marketers*, HOOTSUITE (Dec. 20, 2019), <https://blog.hootsuite.com/facebook-demographics/>.

189. Maryam Mohsin, *10 Instagram Stats Every Marketer Should Know in 2020 [Infographic]*, OBERLO (Jul. 6, 2020), <https://www.oberlo.com/blog/instagram-stats-every-marketer-should-know>.

of the user's political beliefs and are made within the context of an argument. Courts should view this statement as nonactionable hyperbole. On the other hand, the statement, "@TwitterUser2 is a far-right extremist who participated in the U.S. Capitol insurrection" is a more serious accusation. This statement could cause @TwitterUser2 to lose his job, face social ostracization, and even be investigated by the government if it believes this statement may be true.¹⁹⁰ Allegations of either a person's membership in an extremist group, or their participation in an attempted coup, would both be specific enough to be proven either true or false; therefore, a court should look at this statement as potentially actionable defamation.

Parsing the line between fact and opinion, particularly within the social media context, can be difficult given the conversational nature of these platforms. Although the aforementioned analysis may increase the liability social media users face when publishing content, it will provide an effective remedy to individuals harmed by false and damaging statements. Furthermore, other defenses like truthfulness can serve as backstops against this increased defamation liability. This analysis will ensure users are able to freely exercise their First Amendment rights on the Internet without impermissibly chilling their speech.

B. Influencers: The New Limited-Purpose Public Figures

One way that courts may be able to counterbalance the heightened liability placed on social media users for damaging, false statements is to increase the requisite level of intent for this defamation. Increasingly, established consumer brands and companies turn to social media influencers and viral content publishers, rather than to celebrities, to create sponsorships and build content partnerships. These influencers publish a wide range of content, including clothing reviews, lifestyle marketing, travel blogs, and even "day in the life" relatable posts. In turn, other social media users follow, subscribe, and repost this content, expanding its reach and influence. These influencers' voluntary involvement in public life, and their engagement in a "pattern of activity destined to invite attention and comment," make them more like limited-purpose or even general-purpose public figures than

190. See AJ Willingham and Carma Hassan, *People at the U.S. Capitol Riot Are Being Identified and Losing Their Jobs*, CNN (Jan. 9, 2021), <https://www.cnn.com/2021/01/07/us/capitol-riots-people-fired-jobs-trnd/index.html>.

private individuals.¹⁹¹ Consequently, statements made about this emerging group of public figures would be held to the more stringent actual malice liability standard.

Whether users are public or private individuals for liability purposes should not necessarily be determined by only looking at their social media profiles' follower, like, or subscriber totals.¹⁹² Directing courts to look at numbers alone is not helpful: for instance, users with relatively fewer profile followers may still be considered influential in their respective fields (such as within the LGBTQ+ community or specific political circles). Although Twitter's authentication of an account of public interest through blue check verification may be helpful in separating "online trolls" from legitimate accounts, it is not determinative of social status or notoriety.¹⁹³ Courts should instead look to the "three R's" used to measure impact in the marketing industry: reach, resonance, and reaction.¹⁹⁴ Reach is defined as the number of people who have seen the content or profile. Resonance is the ability for the individual to influence consumption or other habits. Finally, reaction is the return on investment for the individual, measured by changes to follower or subscriber count, product sales, or even endorsement requests.

Once social media influencers and viral stars are considered limited-purpose public figures using the "three R's," any allegedly defamatory statements published about them that relate to an "issue of public concern" should be assessed under the actual malice standard. This means that users would only be liable for statements made with knowledge of their falsity or with reckless disregard for the truth. Thus, the standard would protect expressions of free speech and allow the "breathing space" necessary for open and robust public dialogue.¹⁹⁵

Nonetheless, the actual malice standard would not apply to those involuntarily thrust into the spotlight, such as social media users who are attacked by "online mobs" for their opinions and who otherwise do not seek public attention. As the Court has recognized, "instances of

191. Batza, *supra* note 130, at 451–52.

192. *But see id.* at 465 (arguing that Twitter users "do not have equal voices," so courts should look to follower count to determine the "command of the channel that the plaintiff has").

193. *About Verified Accounts*, TWITTER, <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts> (last accessed Mar. 2, 2021).

194. *Our Recipe for Successful Marketing and Media Strategies*, NIELSEN, <https://www.nielsen.com/us/en/solutions/reach-resonance-reaction/> (last accessed Mar. 2, 2021).

195. *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 272 (1964).

truly involuntary public figures must be exceedingly rare.”¹⁹⁶ Social media, however, is an environment that supports not only publishing on matters of public concern, but also treating the platform like a diary for personal anecdotes intended for a smaller social circle’s view. If one of these posts is incidentally shared widely or temporarily featured in “Twitter Top Trends,” it should not automatically mean that the user is no longer considered a private individual for purposes of defamation analysis. Consequently, social media users who take the opportunity to defame, dox, or otherwise harm these users could still be held liable under a negligence standard for posting false statements of fact. Only when a user voluntarily enters into the spotlight—such as by penning an op-ed, accepting brand sponsorships to raise profile visibility, or by speaking to the media on unrelated topics—should she be considered a limited-purpose public figure.

C. Equalized Treatment of Individual Platforms

Courts have often treated the tone of social media statements differently, depending on the platform on which the statements were published. For example, bulletin boards and chat rooms are treated as a source of “casual, emotive, and imprecise speech.”¹⁹⁷ In *John Doe No. 1 v. Cahill*, the court alluded to this differential treatment of sources:

Ranked in terms of reliability, there is a spectrum of sources on the [I]nternet. For example, chat rooms and blogs are generally not as reliable as the *Wall Street Journal Online*. Blogs and chat rooms tend to be vehicles for the expression of opinions; by their very nature, they are not a source of facts or data upon which a reasonable person would rely.¹⁹⁸

In contrast to their treatments of chat rooms, online bulletin boards, and blogs, courts are much more willing to treat online review sites like Yelp as serious sources of factual information. This is likely due to the way these platforms are used by consumers to evaluate products and businesses in order to make purchasing decisions. Because users treat these platforms as more reliable, courts are less forgiving of false statements that have the risk of incurring real harm on companies.

Courts are less consistent, however, when it comes to the treatment of statements made on Twitter. For instance, the *Jacobus* court

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

197. Sandals Resorts Int’l Ltd. v. Google, 925 N.Y.S.2d 407, 415 (N.Y. App. Div. 2011) (citing O’Brien, *supra* note 101, at 2774–75).

198. John Doe No. 1 v. Cahill, 884 A.2d 451, 465 (Del. 2005).

considered tweets to be “per se hyperbole,”¹⁹⁹ while the *Unsworth* court determined tweets can still contain actionable facts.²⁰⁰ Although both tweets were published over the course of very public arguments and contained casual, emotive language, they were treated differently, but they should not have been. Within the context of Twitter, courts should be careful not to dismiss tweets as merely exaggerated expressions or opinions. Now that Twitter has become a global source of breaking news,²⁰¹ and a majority of Americans rely on it as their primary news source,²⁰² tweets are being taken seriously. False Twitter statements could therefore have serious ramifications for individuals’ reputations, employment opportunities, and finances. Social media posts are easily accessible through search engine results, which create a permanent online catalog of an individual’s persona. Courts should focus on the language contained within a social media post, regardless of its platform, to determine whether a statement at issue is actionable.

D. *Reevaluating Jacobus v. Trump*

If the *Jacobus v. Trump* court conducted its defamation analysis using the approach recommended here, it might have found an actionable false statement of fact, rather than incorrectly dismissing all social media speech as nonactionable hyperbole. The statements made on Twitter accused Jacobus of having “begged us for a job,” and claimed that the Trump campaign “[t]urned her down twice and she went hostile.”²⁰³ In truth, Jacobus met with campaign officials twice before letting them know that she would not be interested in working with the campaign. The tweets also claimed that Jacobus was a “real dummy,” “really dumb,” and a “[m]ajor loser, zero credibility.”²⁰⁴

The court dismissed these statements as part of a “petty quarrel,” the “familiar back-and-forth between a political commentator and the

199. *Jacobus v. Trump*, 51 N.Y.S.3d 330, 339 (N.Y. Sup. Ct. 2017).

200. *Unsworth v. Musk*, No. 2:18-cv-08048-SVW-JC, 2019 U.S. Dist. LEXIS 167626 at *1 (C.D. Cal. 2019).

201. See, e.g., Donald J. Trump (@realDonaldTrump), TWITTER (Mar. 13, 2018, 8:44 AM), <https://web.archive.org/web/20210104030911/https://twitter.com/realDonaldTrump/status/973540316656623616>; Donald J. Trump (@realDonaldTrump), TWITTER (Jul. 26, 2017, 9:04 AM), <https://web.archive.org/web/20210108045647/https://twitter.com/realDonaldTrump/status/890196164313833472>; Donald Trump (@realDonaldTrump), TWITTER (Jun. 7, 2017, 7:44 AM), <https://web.archive.org/web/20210107000216/https://twitter.com/realDonaldTrump/status/872419018799550464>.

202. Grieco & Shearer, *supra* note 1.

203. *Jacobus v. Trump*, 51 N.Y.S.3d 330, 334 (N.Y. Sup. Ct. 2017).

204. *Id.*

subject of her criticism.”²⁰⁵ This dismissal was premature, because in fact these statements were not common rebuttals of critical political commentary. Instead, they were spiteful responses meant to undermine Jacobus’s credibility. The court also excused the statements as part of Trump’s regular skewering of his opponents and critics. Under the analysis set forth in this Note, the statements could instead be viewed as actionable false statements of fact. The statement “begged us for a job” posits that Jacobus aggressively approached the campaign for a position. Additionally, the statement that the campaign “[t]urned her down twice” could be demonstrably false, proven through e-mail and text communications documenting Jacobus’s rejection of the position. Although the court viewed these statements as boisterous exaggerations, they portrayed Jacobus as someone desperate to work for the Trump campaign. This portrayal undermined her credibility as a political commentator and critic of Trump, and it cost her future job opportunities.

Next, the court should have determined whether Jacobus was a public or private figure. This would have been a relatively easy analysis: Jacobus was a political commentator on television news channels and other media outlets, including CNN.²⁰⁶ She inserted herself into the Trump campaign coverage and was at least a limited-purpose public figure for the purposes of this controversy. Consequently, Jacobus would have had to prove that Trump made these statements with knowledge of their falsity or with reckless disregard of the truth. Because Jacobus exchanged texts with campaign officials when she turned down the position, it may have been possible to prove Trump’s knowledge of the situation. This would have been a matter of examining the evidence, which the court did not adequately do.

Finally, the court erred when it focused on the platform’s nature in its analysis of the statement’s context. The court placed too much weight in its view that social media was a less credible source of information.²⁰⁷ Twitter may be a platform for casual communications, but it is just as equally a platform for breaking news and serious information. The court should have evaluated these statements just as carefully as if they were made in a *New York Times* op-ed.

205. *Id.* at 342.

206. *Id.* at 334.

207. *Id.* at 339.

E. Response to Anticipated Critiques

This approach raises several valid concerns. First, critics may contend that due to the democratized access to social media platforms, users unfamiliar with the risks of publishing false statements may find themselves facing litigation over tweets and posts they assumed were harmless missives. Although it is true that many social media users may not fully understand the power of their words broadcast across the Internet, they certainly would recognize the risk of publishing the same content on more traditional platforms like newspapers. Social media statements have just the same power to damage a person's reputation and cause economic losses; consequently, they should be treated as seriously as traditionally-published statements under defamation analysis. Second, critics might point to the opening of floodgates of litigation if individuals sued social media users for every careless statement posted. True, individuals may be emboldened to bring claims against social media users who have posted false statements of fact if more courts award damages for social media defamation. However, these claims will still have to survive the pleading and motion to dismiss stages, so frivolous defamation litigation will not withstand scrutiny. Courts will also look to the context of statements made when assessing these social media defamation claims, distinguishing between serious, provable statements and more casual, exaggerated speech that falls under protected opinion. Finally, while this framework may be criticized as an imprecise balancing test, such is the nature of defamation litigation generally. One must determine from the unique facts of each case whether the speech rises to the level of a false statement of fact.

CONCLUSION

Defamation cases involving social media speech present more gray areas than cases involving traditional media outlets. Users are more likely to remain anonymous; casual and imprecise speech may be presented as opinion rather than fact; the fast-paced nature of online speech often means statements are posted without much second thought. But because the Internet is a global communications platform with a permanent memory, this speech may be just as damaging, if not more so, as a news article or broadcast to a person's reputation. Courts must weigh the benefits of an expansive marketplace of ideas on the Internet with the consequences of widespread false and defamatory speech. Furthermore, they must reckon with the development of the

social media influencer and the elevation of once-private individuals to a status more like limited-purpose public figures. They should do so by increasing the level of intent required to hold a publisher accountable for its false assertions of fact to an actual malice standard. As people increasingly view social media platforms as reliable news sources, so too must they face responsibility for the content they publish on these platforms. If courts treat all social media speech as opinion, shielded from liability, disinformation and baseless personal smears will continue to run rampant on the Internet without any effective guardrails.