*Harvard Journal of Law & Technology*
Volume 31, Number 1 Fall 2017

# THE CONSTITUTIONALITY OF CRIMINALIZING FALSE SPEECH MADE ON SOCIAL NETWORKING SITES IN A POST-*ALVAREZ*, SOCIAL MEDIA-OBSESSED WORLD

*Louis W. Tompros, Richard A. Crudo, Alexis Pfeiffer, Rahel Boghossian**

TABLE OF CONTENTS



---

* Louis W. Tompros is a lecturer on law at Harvard Law School and a partner in the Boston office of WilmerHale LLP. Richard A. Crudo is a former senior associate in the Washington, D.C. office of WilmerHale. All research and drafting of this article occurred while Mr. Crudo was employed at WilmerHale. Alexis Pfeiffer is an associate in the Palo Alto office of WilmerHale. Rahel Boghossian is a former summer associate in the Washington, D.C. office of WilmerHale.

## I. INTRODUCTION

> [A] NATION THAT IS AFRAID TO LET ITS PEOPLE JUDGE THE TRUTH AND FALSEHOOD IN AN OPEN MARKET IS A NATION THAT IS AFRAID OF ITS PEOPLE.
>
> — JOHN F. KENNEDY[1]
>
> FREE SPEECH HAS REMAINED A QUINTESSENTIAL AMERICAN IDEAL, EVEN AS OUR SOCIETY HAS MOVED FROM THE INK QUILL TO THE TOUCH SCREEN.
>
> — MARVIN AMMORI[2]

The emergence of social media led to profound changes in the way we interact with technology and each other. Every day — often without thinking — we use social media platforms for myriad purposes, including to keep family and friends apprised of developments in our lives, to reconnect with long-lost friends, to debate contemporary social and political issues, to conduct business, and even to find romance. It is unsurprising, therefore, that social media established itself as a worldwide phenomenon. According to current estimates, there are nearly 2.8 billion users of social media worldwide, and that number is expected to increase

1. John F. Kennedy, Remarks on the 20th Anniversary of the Voice of America (Feb. 26, 1962), http://www.presidency.ucsb.edu/ws/?pid=9075 [https://perma.cc/Z4CJ-BH72].

2. Marvin Ammori, *Should Copyright Be Allowed to Override Speech Rights?*, THE ATLANTIC (Dec. 15, 2011), http://www.theatlantic.com/politics/archive/2011/12/should-copyright-be-allowed-to-override-speech-rights/249910/ [https://perma.cc/JMS6-DSDN].

dramatically over the next several years.[3] There are now hundreds of thousands of messages and posts on social media websites and mobile apps occurring every minute.[4] As several Supreme Court Justices recently observed, social media is "embedded in our culture," and there is perhaps no other forum in history that is so accessible and in which speech is so prolific.[5]

But "with the advent of social media and modern digital communication there is great opportunity for individuals to perpetuate mischief that can result in falsehoods."[6] The fake news epidemic that recently dominated the headlines provides an obvious example of such falsehoods, but there are many others. Hyperbole, embellishment, practical jokes, rumors, catfishing,[7] and even malicious lies and threats are not uncommon on social media. Indeed, it is well documented that social media led to a more cavalier attitude about the truth; social media's veil of actual (or perceived) anonymity allows subscribers to more aggressively spread falsehoods.[8] To be sure, many of these lies are innocuous enough. It is not uncommon, for example, for users to exaggerate about their lives to improve their social status, or for a person to lie about his height or weight in his online profile in an effort to appear more desirable to would-be suitors.[9] These lies are often calculated (perhaps subconsciously) to subvert one's real-life persona with an upgraded cyber persona. But some lies are much more injurious.

---

3. Simon Kemp, *Digital in 2017: Global Overview*, WE ARE SOCIAL (Jan. 24, 2017), https://wearesocial.com/blog/2017/01/digital-in-2017-global-overview [https://perma.cc/W68C-S9R2].

4. *See* Jonathan Shaw, *Why "Big Data" Is a Big Deal*, HARV. MAG., Mar.-Apr. 2014 (quoting Gary King, Director of the Institute for Quantitative Social Science at Harvard University); *see also* Hillary Rodham Clinton, Sec'y of State, Remarks on Internet Freedom (Jan. 21, 2010), *available at* http://foreignpolicy.com/2010/01/21/internet-freedom [https://perma.cc/SA4Y-WB6V] (observing in 2010 that "[t]here are more ways to spread more ideas to more people than in any moment in history").

5. Transcript of Oral Argument at 32, Packingham v. North Carolina, 137 S. Ct. 1730 (2017) (No. 15-1194), https://www.supremecourt.gov/oral_arguments/argument_transcripts/2016/15-1194_0861.pdf; *see also id.* at 28 (noting that communication via social media is "greater than the communication you could ever [have], even in the paradigm of public square").

6. *Ex parte* Maddison, 518 S.W.3d 630, 634 (Tex. App. 2017) (citing trial court's opinion).

7. *Catfish,* MERRIAM-WEBSTER (2017) ("a person who sets up a false personal profile on a social networking site for fraudulent or deceptive purposes").

8. *See* Aditi Gupta & Ponnurangam Kumaraguru, *Credibility Ranking of Tweets During High Impact Events*, PROC. 1ST WORKSHOP ON PRIV. & SEC. ONLINE SOC. MEDIA (2012), https://dl.acm.org/citation.cfm?id=2185356 (last visited Oct. 24, 2017); Paul Grabowicz, *Tutorial: The Transition to Digital Journalism*, KDMC BERKELEY (Mar. 30, 2014), https://multimedia.journalism.berkeley.edu/tutorials/digital-transform/ [https://perma.cc/4PKC-QNKD].

9. *See generally, e.g.*, MARY AIKEN, THE CYBER EFFECT 172–74 (2016) (discussing "the obsessive interest among teens" in manipulating and curating selfies and their online profiles in an effort to portray their best "cyber self"); *id.* at 217 (noting that, "[w]hile some individuals may use cyber-dating to experiment with new selves, new behaviors, or a new gender, there are other people who just like to lie about who they are — and trick strangers").

We have seen several recent examples in which social media users publish false information about emergencies and natural catastrophes. This effect was perhaps most prevalent in the 2013 Boston Marathon bombings, when news outlets relied on social media postings to falsely identify innocent people as the perpetrators, mistakenly report that the perpetrators were arrested, and incorrectly claim that additional explosive devices were discovered.[10] The effect was also noticeable in online reports of other terrorist attacks, mass shootings, earthquakes, hurricanes, and other emergencies.[11] These false reports are significant, as social media has now established primacy over traditional news outlets like cable and radio, at least for the cyber savvy.[12] Indeed, more than 60% of Americans now get their news from social media websites and apps like Facebook and Twitter.[13] False reports of emergencies are therefore likely to be read and rebroadcast by many people, leading to their uncontrolled propagation through cyberspace and, potentially, mass hysteria. Arguably, false reports of emergencies and natural catastrophes are, in some instances, the digital equivalent of yelling "fire!" in a crowded theater.

Traditionally, such speech was thought to fall outside the realm of First Amendment protection. But recent Supreme Court authority may require us to revisit that conclusion. In 2012, the Supreme Court issued its *United States v. Alvarez*[14] decision, in which the Court struck down the Stolen Valor Act of 2005, which made it a crime to falsely claim receipt of military decorations or medals. In so holding, the Court established a First Amendment right, in some circumstances, to lie. Thus, *Alvarez* provides powerful support for the notion that some lies spread on social media may be protected. Additionally, the very nature of the internet limits the scope of the harm caused by lies made on social media. Although lies may be rebroadcast many times in a matter of minutes, social media subscribers are able to easily vet and rebuff falsehoods with just a click of a mouse. This self-correcting — or, more accurately, crowd-correcting — mechanism often allows social media to strike down lies before they travel too deeply into cyberspace.[15] Thus, the con-

10. For a summary of all the false reports made on Twitter during the Boston Marathon bombings, see Christina Reinwald, *What Twitter Got Wrong During the Week Following Last Year's Boston Marathon*, BOSTON.COM (Apr. 18, 2014), https://www.boston.com/news/local-news/2014/04/18/what-twitter-got-wrong-during-the-week-following-last-years-boston-marathon (last visited October 24, 2017).

11. *Id.*

12. Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RESEARCH CENTER 1, 8 (May 26, 2016), http://www.journalism.org/files/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf [https://perma.cc/8QCG-9KGT].

13. *Id.* at 2.

14. 567 U.S. 709 (2012).

15. *Digital Wildfires in a Hyperconnected World*, WORLD ECON. FORUM (2013), http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/ [https://perma.cc/EVR8-T7H4] (describing propagation of falsehoods on social media,

cern that yelling "fire!" may lead to significant and widespread harm may be far less salient in cyberspace than in a crowded theater.

Many states have false reporting statutes that impose criminal liability on those who engage in false speech related to emergencies or natural catastrophes, regardless of the medium used to communicate the speech. But New York's false reporting statute is perhaps the broadest, and therefore the most likely to be susceptible to a First Amendment challenge. The statute proscribes circulating reports of emergencies or natural catastrophes that the speaker knows are false or baseless and that are "not unlikely" to cause "public alarm or inconvenience."[16] While the statute requires knowledge that the statement is false or baseless, it does not require knowledge or intent with respect to the ensuing public alarm or inconvenience.[17] Additionally, the statute permits liability based on a tenuous nexus to the underlying harm, requiring only that public alarm or inconvenience be not unlikely.[18] The New York statute withstood a pre-*Alvarez* First Amendment challenge, but it is unclear whether the statute would survive after *Alvarez*, specifically as applied to false speech on social media, where there may exist less restrictive alternatives to avoiding the harm imposed by spreading false speech.

The New York statute's breadth makes it an interesting model for examining this issue. In particular, this Article analyzes the viability of a First Amendment challenge to the New York false reporting statute as applied to false speech on social media. The Article begins by describing social media, generally, and its impact on the way that we consume and disseminate news of high-profile events. Next, the Article examines how and why lies spread through social media, and describes the crowd-correcting mechanism that often counteracts the widespread dissemination of such lies. The Article then analyzes New York's statute and its theoretical underpinnings, as well as First Amendment challenges thereto. The Article next sets forth a First Amendment framework for analyzing false speech, which culminates in an analysis of *Alvarez*. Finally, the Article applies that framework to assess the viability of a First Amendment challenge to New York's statute as applied to false speech made on social media. The analysis is grounded in the real-life example of a teenager who suggested on Twitter that his town was going to have a deadly "purge," based on the recent horror films of the same name.

---

crowd-correcting mechanisms, and potential for causing panic); Gerry Shih, *During Hurricane Sandy, Twitter Proves a Lifeline Despite Pranksters Like @ComfortablySmug*, HUFFINGTON POST (Oct. 31, 2012), http://www.huffingtonpost.com/2012/10/31/hurricane-sandy-twitter-comfortablysmug_n_2047754.html [https://perma.cc/FQ9T-TY4S].

16. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).

17. *Id.*

18. *Id.*

## II. THE SOCIAL MEDIA REVOLUTION

### *A. What Is Social Media?*

Shortly after the advent of the modern internet in the early 1990s, we beheld a cultural and technological revolution involving social media — "a group of Internet-based applications that . . . allow the creation and exchange of User Generated Content."[19] Social media transformed the way we interact with technology as well as how we engage with others. Today, "socialmedialites"[20] can inform thousands of friends and acquaintances — and oftentimes, total strangers — of their activities, political and social opinions, and impressions with a click of a mouse. Social trends are now dictated by internet "memes" and viral YouTube videos that propagate fluidly and swiftly through cyberspace.[21] "There are now a billion social-media posts every two days . . . which represents the largest increase in the capacity of the human race to express itself at any time in the history of the world."[22] It is no wonder, then, that we now have an annual (unofficial) holiday, "Social Media Day," to help us "highlight the ways digital culture has revolutionized how we communicate."[23]

As of January 2017 there were between 2.4 and 2.8 billion active social media users in the world.[24] That number is expected to increase to

---

19. Andreas M. Kaplan & Michael Haenlein, *Users of the World, Unite! The Challenges and Opportunities of Social Media*, 53 BUS. HORIZONS 59, 61 (2010).

20. Urban Dictionary defines "socialmedialite" as "a person who participates in social media, spends a significant amount of time promoting themselves at fashionable events and promoting themselves through social media channels; A social media darling." *Socialmedialite*, URBAN DICTIONARY (May 20, 2014), http://www.urbandictionary.com/define.php?term=Socialmedialite [https://perma.cc/8LYW-JHNG].

21. *See generally* LINDA K. BÖRZSEI, MAKES A MEME INSTEAD: A CONCISE HISTORY OF INTERNET MEMES (2013) (investigating the ontology, history, and evolution of the internet meme — i.e., content that spreads online from user to user and changes along the way), *available at* https://works.bepress.com/linda_borzsei/2/ [https://perma.cc/ZX7Q-X8CS]; HENRY JENKINS ET AL., IF IT DOESN'T SPREAD, IT'S DEAD: CREATING VALUE IN A SPREADABLE MARKETPLACE 1, 2 (2008) http://convergenceculture.org/research/Spreadability_doublesidedprint_final_063009.pdf [https://perma.cc/N7QZ-TW4N] (analyzing examples of internet "memes" and "viruses" and how they have evolved, and proposing an alternative model involving "spreadable media" in shaping the circulation of media content); *see also* Jure Leskovec et al., *Meme-Tracking and the Dynamics of the News Cycle*, PROC. 15TH INT'L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 497 (2009) (discussing a framework for tracking short, distinctive phrases that travel through online text and observing a lag of 2.5 hours between the peaks of attention to a phrase in the news media and in blogs).

22. Shaw, *supra* note 4.

23. Lulu Chang, *Today, We're Celebrating Social Media Day, Otherwise Known as Thursday*, DIGITAL TRENDS (June 30, 2016), http://www.digitaltrends.com/social-media/social-media-day/ [https://perma.cc/AG86-FMJT].

24. Kemp, *supra* note 3.

nearly 3 billion by 2020.[25] There are now hundreds of social media platforms available, including Facebook, Twitter, YouTube, LinkedIn, Google+, Reddit, Pinterest, and Instagram. Each of these platforms allows users to interact with one another by sharing text, images, and/or videos of interests, hobbies, and news.

### *B. Social Media as a Reliable News Source or a Gossip Platform?*

Although much of the content of social media has been categorized as "pointless babble"[26] — e.g., breakfast-cereal updates, interesting new links, and music recommendations[27] — social media's use transcends the banal observations and musings of its constituency. Studies show that 85% of topics discussed on social media platforms such as Twitter are related to events in the news.[28] In fact, a 2016 Pew Research Center study found that 62% of American adults get their news through social media, representing an increase from 49% in 2012.[29] And oftentimes government authorities and the Associated Press will take to social media before officially publishing statements or articles related to breaking news in traditional outlets.[30] For example, reports that Osama Bin Laden was killed in 2011 broke on Twitter hours before President Obama addressed the nation with the news.[31] Thus, social media serves not only as a social network, but also as a vehicle for delivering the latest news.

---

25. *Number of Social Media Users Worldwide from 2010 to 2020 (in billions)*, STATISTA, https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [https://perma.cc/XEG5-C9Q8].

26. PEAR ANALYTICS, TWITTER STUDY 4–5 (2009), http://web.archive.org/web/20110715062407/www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf (last visited Dec. 20, 2017).

27. Steven Johnson, *How Twitter Will Change the Way We Live*, TIME, June 5, 2009, http://content.time.com/time/magazine/article/0,9171,1902818,00.html (last visited Dec. 20, 2017).

28. Haewook Kwak et al., *What Is Twitter, a Social Network or a News Media?*, 19TH INT'L CONF. ON WORLD WIDE WEB 1, 10 (2010).

29. Gottfried and Shearer, *supra* note 12.

30. *See* EDWARD F. DAVIS III ET AL., HARV. KENNEDY SCH., SOCIAL MEDIA AND POLICE LEADERSHIP: LESSONS FROM BOSTON 3–4 (Mar. 2014), https://www.ncjrs.gov/pdffiles1/nij/244760.pdf [https://perma.cc/7DUS-F5DG?type=image] (noting that Boston police focused on using social media to "push[] accurate and complete information to the public" as soon as possible during the 2013 Boston Marathon bombings); Joe Coscarelli, *Associated Press Staff Scolded for Tweeting Too Quickly About OWS Arrests*, N.Y. MAG., Nov. 16, 2011, http://nymag.com/daily/intelligencer/2011/11/ap-staff-scolded-for-tweeting-about-ows-arrests.html [https://perma.cc/7DUS-F5DG?type=image] (discussing Associated Press's missive to its employees admonishing them not to "break news that [has not been] published, no matter the format" after employees preemptively tweeted their arrests during the 2011 Occupy Wall Street protests).

31. Brian Stelter, *How the Bin Laden Announcement Leaked Out*, N.Y. TIMES (May 1, 2011, 11:28 PM), http://mediadecoder.blogs.nytimes.com/2011/05/01/how-the-osama-announcement-leaked-out/ [https://perma.cc/DT4A-9S4E].

1. Social Media's Use During High-Profile Events and Crises

Social media has also proven popular for communicating in real time about emergency crises.[32] "The immediacy, ease of access, and widespread use of social media channels like Facebook, LinkedIn and Twitter make these digital platforms a hot-bed for breaking news."[33] For example, close to 35% of tweets sent as Hurricane Sandy made landfall and pummeled its way up the East Coast in October 2012 were news related.[34] Social media likewise played an important role as a source of information during the 2007 fires that raged across Southern California; the 2008 New England ice storm that wiped out power for 400,000 homes and businesses in the region; the 2008 Sichuan earthquake, which killed almost 70,000 people; and the 2008 cyclone in Myanmar, which caused major destruction and nearly 150,000 fatalities.[35] Additionally, more than 27 million tweets were sent during the April 2013 Boston Marathon bombings, when an intense three-day manhunt ensued after twin explosions at the Boston Marathon killed three people and injured 264 others.[36] But social media users are not just passive recipients of the news in such circumstances; they are often *creators* of the news.

Indeed, social media content frequently serves as source material for news media reports.[37] This was apparent during the Boston Marathon bombings, when news media relied on tweets to falsely identify innocent people as the perpetrators, mistakenly report that the perpetrators had been arrested, and incorrectly claim that additional explosive devices were discovered.[38] Other examples demonstrate that the news media's

---

32. *See* Michel Martin, *Tell Me More: Why Some Spread Misinformation in Disasters*, NPR (Nov. 2, 2012), http://www.npr.org/2012/11/02/164178388/why-some-spread-misinformation-in-disasters [https://perma.cc/4ED7-PYAV] (describing the "good, the bad, and the ugly of social media" during Hurricane Sandy in October 2012).

33. Mostafa Razzak, *Breaking News with Social Media*, INTERNET MARKETING ASSOCIATION (Feb. 1, 2016), https://imanetwork.org/industry-news/breaking-news-with-social-media/ [https://perma.cc/A3VF-DMRQ].

34. *See* Emily Guskin & Paul Hitlin, *Hurricane Sandy and Twitter*, PEW RES. CTR. (Nov. 6, 2012), http://www.journalism.org/2012/11/06/hurricane-sandy-and-twitter/ [https://perma.cc/MW8K-84CH].

35. Alexander Mills et al., *Web 2.0 Emergency Applications: How Useful Can Twitter Be for Emergency Response?* 5 J. INFO. PRIV. & SEC. 3, 14–16 (2009). For a summary of related research on social media's use during news events, see Carlos Castillo et al., *Predicting Information Credibility in Time-Sensitive Social Media*, 23 INTERNET RES. 560 (2012); Gupta & Kumaraguru, *supra* note 8; and Grabowicz, *supra* note 8.

36. *The Year in Twitter: Top Milestones of 2013*, MASHABLE, http://mashable.com/2013/12/12/twitter-2013/#1TmiWo10hgqV 407550881433792512 [https://perma.cc/NJZ7-L27K]; *Boston Marathon Bombing*, WIKIPEDIA, https://en.wikipedia.org/wiki/Boston_Marathon_bombing [https://perma.cc/BNQ2-6C7H].

37. Brooke Gladstone & Bob Garfield, *On the Media: Coverage of Boston, Uncovered Reporting and More*, NPR (Apr. 19, 2013), http://www.onthemedia.org/story/287989-coverage-of-the-boston-bombing-undercover-reporting-and-more/ [https://perma.cc/8NRS-P3UP] (noting that reports on police scanners parroted false tweets during the Boston Marathon bombings).

38. *See* Reinwald, *supra* note 10.

increased and unquestioned reliance on social media, although disturbing in some instances, has dramatically changed the landscape of journalism.[39] In 2014, CNN announced that it had partnered with analytics firm Dataminr to develop a tool that scans Twitter for newsworthy trends and alerts journalists to breaking stories,[40] taking advantage of the "democratization of headline news and emergent social behavior such as crowd-sourcing"[41] that social media helped effectuate. Indeed, social media is excellent for "providing information not covered on radio and television, such as details and first-hand accounts within moments of an event, anywhere in the world. There is no other medium that can compete with [social media] in that arena."[42] Thus, in some ways, social media has become a de facto emergency broadcast channel.[43] The demarcation between *social* media and *news* media is now blurred — social media has made journalists of us all, whether we like it or not.[44]

But unlike reports from journalists, social media posts are typically not vetted for accuracy or veracity.[45] Due to the often "anonymous and

---

39. *See generally* Adam Cohen, *The Media that Need Citizens: The First Amendment and the Fifth Estate*, 85 S. CAL. L. REV. 1 (2011). The issue of whether journalists may legally and ethically rely on social media as a source of news is interesting, but beyond the scope of this Article. For a discussion of how social media has led to "ambient journalism" and how awareness systems impact journalism, see Alfred Hermida, *Twittering the News: The Emergence of Ambient Journalism*, 4 JOURNALISM PRAC. 297 (2010). For a discussion of the news media's reliance on "iReporting" and legal liability therefor, see Virginia A. Fitt, *Crowdsourcing the News: News Organization Liability for iReporters*, 37 WM. MITCHELL L. REV. 1839 (2011), and Kimberly Chow, Note, *Handle with Care: The Evolving Actual Malice Standard and Why Journalists Should Think Twice Before Relying on Internet Sources*, 3 N.Y.U. J. INTELL. PROP. & ENT. L. 53 (2014).

40. Jason Abbruzzese, *CNN Doubles Down on Twitter-Based Reporting with Dataminr Deal*, MASHABLE (Jan. 29, 2014), http://mashable.com/2014/01/29/cnn-doubles-down-on-twitter-based-reporting-with-dataminr-partnership/ [https://perma.cc/9PSB-DRW6].

41. Mills et al., *supra* note 35, at 6; *see also* Fitt, *supra* note 39.

42. Mills et al., *supra* note 35, at 21.

43. Jeff Roberts, *Tweeting Fake News in a Crisis — Illegal or Just Immoral?*, GIGAOM (Oct. 30, 2012, 1:17 PM), https://gigaom.com/2012/10/30/tweeting-fake-news-in-a-crisis-illegal-or-just-immoral/ [https://perma.cc/Z5KX-SWK8].

44. Indeed, the United States Court of Appeals for the Ninth Circuit issued a 2014 decision that further blurs this line, holding that bloggers — i.e., authors of websites that maintain an ongoing chronicle of information and commentary — have some of the same First Amendment rights as bona fide journalists. *See* Obsidian Fin. Grp. v. Cox, 740 F.3d 1284 (9th Cir. 2014).

45. Social media companies are struggling to find a balance between curbing false reports on their sites and protecting expression. For example, Facebook Chief Executive Mark Zuckerberg stated that Facebook will not try to separate fact from fiction because "[w]e must be extremely cautious about becoming arbiters of truth ourselves." Deepa Seetharaman, Jack Nicas & Nathan Olivarez-Giles, *Social-Media Companies Forced to Confront Misinformation and Harassment: Sites Struggle to Find a Balance Between Being Havens for Misinformation and Censors of Free Speech*, WALL ST. J. (Nov. 15, 2016), https://www.wsj.com/articles/social-media-companies-forced-to-confront-misinformation-and-harassment-1479218402 (last visited Dec. 20, 2017). But after facing intense scrutiny for the spread of fake news and misinformation on its platform during the 2016 presidential election, Facebook decided to allow fact-checkers to verify links shared on Facebook, to tweak the News Feed ranking algorithm, and to create easier ways for users to flag fake news. Craig Silverman, *Facebook is Turning to Fact-Checkers to Fight Fake News*, BUZZFEED (Dec. 15, 2016, 11:59 AM),

unmonitored nature of the Internet, a lot of content generated on [social media] maybe [sic] incredible."[46] And even if not technically false, social media posts can be misleading given the difficulty of providing essential details and context in just a limited number of words. It is not surprising, therefore, that unsubstantiated reports about newsworthy events that turn out to be false or inaccurate are widely circulated via social media. A recent study found that only 17% of content on Twitter related to any contemporaneously occurring emergency event is credible.[47] Another study analyzed 7.8 million tweets related to the Boston Marathon bombings and discovered that 29% of the most viral content comprised rumors and false reports.[48]

The false tweets during the Boston Marathon bombings represent just the tip of the iceberg. For example, in the aftermath of the November 2015 Paris terrorist attacks, which resulted in the deaths of 129 people, several social media sites were flooded with rumors and misinformation regarding facts surrounding the tragedy.[49] In 2013, news media exploded with reports that President Obama had been injured in a bombing at the White House after the Associated Press's Twitter account was hacked.[50] Similarly, in 2011, false reports that President Obama had been killed by an assailant's bullet while campaigning in Iowa issued from Fox News' hacked social media account before being taken down.[51] Further, in 2014, tweets surfaced reporting that Malaysia Airlines Flight 370, which is thought to have crashed into the Indian Ocean shortly after it departed from Kuala Lumpur in March 2014, safe-

---

https://www.buzzfeed.com/craigsilverman/facebook-and-fact-checkers-fight-fake-news?utm_term=.mfv5elaqn#.yq46Pgkz5 [https://perma.cc/XKT4-3PQ3].

46. Gupta & Kumaraguru, *supra* note 8, at 1.

47. *Id.* at 2.

48. Aditi Gupta et al., *$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter*, eCRIME RESEARCHERS SUMMIT (Sept. 17, 2013), http://ieeexplore.ieee.org/document/6805772/ [https://perma.cc/U2HY-X2K6]; *see also* Paul Hitlin, *False Reporting on the Internet and the Spread of Rumors: Three Case Studies*, 4 GNOVIS J. COMM., CULTURE & TECH. (2004), http://www.gnovisjournal.org/files/Paul-Hitlin-False-Reporting-on-the-Internet.pdf [https://perma.cc/NW94-WWYE] (examining the pre-Twitter spread of rumors online vis-à-vis (1) the 1996 crash of TWA flight 800, (2) the report that former White House special assistant Sidney Blumenthal physically abused his wife, and (3) rumors that the death of former Bill Clinton aide Vince Foster was a murder, not a suicide).

49. Sarah Whitten, *Rumors and Misinformation Circulate on Social Media Following Paris Attacks*, CNBC (Nov. 14, 2015, 3:00 PM), http://www.cnbc.com/2015/11/14/rumors-and-misinformation-circulate-on-social-media-following-paris-attacks.html [https://perma.cc/YD34-NUSY].

50. Rebecca Shapiro, *AP Twitter Account Hacked*, HUFFINGTON POST (Apr. 23, 2013, 1:21 PM), http://www.huffingtonpost.com/2013/04/23/ap-twitter-hacked_n_3140277.html [https://perma.cc/T36E-R9X2].

51. *See* Liz Robbins & Brian Stelter, *Hackers Commandeer a Fox News Twitter Account*, N.Y. TIMES, (July 4, 2011), http://www.nytimes.com/2011/07/05/business/media/05fox.html?pagewanted=all&_r=0 (last visited Dec. 20, 2017).

ly landed in China.[52] And curiously, tweets emanating from a student-run Pennsylvania State University social media account prematurely reported that hall of fame college football coach Joe Paterno died one day before he actually passed away from lung cancer.[53] This report was rebroadcasted by CBS Sports, The Huffington Post, and MSNBC.com before meeting its demise.[54] Finally, after the 2016 presidential election, a post from a little-known right-wing blog erroneously stating that Donald Trump defeated Hillary Clinton in the popular vote appeared atop the Google search results for several election-related queries.[55]

These are just a handful of examples in which false reports of newsworthy events have been made on social media.

### 2. "Digital Wildfire": Why and How Social Media Propagates False Information

The false reports discussed above spread rapidly through cyberspace, like a "digital wildfire."[56] There are at least three explanations for why social media is susceptible to the propagation of this digital wildfire. First, lying on social media is easier and more empowering than lying in real life. Social media allows users to perpetuate lies to a captive audience by portraying personas that the users would never expose or assume in real life, oftentimes protected and encouraged by a veil of anonymity or pseudonymity.[57] Social media helps bring these personas from users' fantasies to reality. And by hiding behind their computer

---

52. Julianne Pepitone, *Social Media Spread False Reports of Safe Landing*, NBC NEWS (Mar. 8, 2014, 9:25 PM), http://www.nbcnews.com/storyline/missing-jet/social-media-spread-false-reports-safe-landing-n48081 [https://perma.cc/CUD9-X5G3].

53. Brian Stelter, *Mistaken Early Report on Paterno Roiled Web*, N.Y. TIMES, (Jan. 22, 2012), http://www.nytimes.com/2012/01/23/business/media/premature-reports-of-joe-paternos-death-roiled-web.html?_r=0 (last visited Dec. 20, 2017). There is no shortage of celebrity death rumors that originate on social media. *See* Amethyst Tate, *Twitter Death Hoaxes of 2012: Morgan Freeman, Bill Cosby, Paris Hilton, Adam Sandler and Others Claimed this Year*, INT'L BUS. TIMES (Sept. 14, 2012, 11:29 AM), http://www.ibtimes.com/twitter-death-hoaxes-2012-morgan-freeman-bill-cosby-paris-hilton-adam-sandler-and-others-claimed [https://perma.cc/2C4X-AFCW] (describing the "death hoax phenomenon" that has claimed, among others, Justin Bieber, Mick Jagger, and Bill Nye).

54. Stelter, *supra* note 53.

55. Seetharaman et al., *supra* note 45.

56. WORLD ECON. FORUM, *supra* 15; Helena Webb et al., *Digital Wildfires: Hyper-Connectivity, Havoc, and a Global Ethos to Govern Social Media*, 45 COMPUTERS & SOC'Y 193 (2015); Helena Webb et al., *Digital Wildfires: Propagation, Verification Regulations and Responsible Innovation*, 34 ACM TRANSACTIONS INFO. SYS. (2016).

57. *See* Paul Bloomfield, *Social Media, Self-Deception, and Self-Respect*, *in* SOCIAL MEDIA AND THE VALUE OF TRUTH 34–35 (Berrin Beasley & Mitchell R. Haney eds., 2013); *see also* Martin, *supra* note 32 (noting that people spread misinformation partly because social media allows them to "explor[e] facets of their personality that they're unable to do offline"); *cf.* Keith Wilcox & Andrew T. Stephen, *Are Close Friends the Enemy? Online Social Networks, Self-Esteem, and Self-Control*, 40 J. CONSUMER RES. 90, 91 (2013) (noting that "social networks allow people to selectively present what they want others to see.").

screens, users insulate themselves from personally and contemporaneously confronting the unpleasant consequences of their falsehoods and from enduring many of the attendant social risks inherent in lying.[58] Social media absolves us of having to uncomfortably look someone in the eye while telling a lie, and, online, every lie seems like a mere fib. Research also shows that it is easier to get away with lying or being someone else when online rather than in real life.[59] This is especially true on social media platforms, where endless series of modified posts could make pinpointing the origin of a lie quite difficult. As a result, most people feel more comfortable lying on social media than in real life.[60] This explains why users intentionally (and sometimes maliciously) spread rumors, tell jokes, and play pranks, but it also explains why people inadvertently or negligently spread false reports with respect to newsworthy events. The blurring between reality and fantasy has led to "a more cavalier attitude to the truth" that has also eroded the distinction between news and entertainment in the post-social media world.[61]

Second, social media allows for instantaneous, real-time provision of news from "ordinary" people who happen to be at the scene of a critical event. Under these circumstances, many users pride themselves on being the first to "break" the news. Indeed, the "drive to be first with the basic facts of a newsworthy development remains embedded in the culture of newsrooms and in the minds of reporters,"[62] and has led to a

---

58. *See* Brian Solis, *The First Amendment of Social Media: Freedom of Tweet*, BRIANSOLIS.COM (Nov. 17, 2010), http://www.briansolis.com/2010/11/the-first-ammendment-of-social-media-freedom-of-tweet/ [https://perma.cc/8QMQ-USB3] ("Inner monologue and filters usually prevent us from uttering words that could haunt us or worse, harm us. Social Media erode these filters enticing us to share in public what might be better shared with discretion. Perhaps our screens shroud us in a protective light.").

59. Bloomfield, *supra* note 57.

60. *People More Likely to Lie on Twitter Than in Real Life, Survey Reveals*, TELEGRAPH (Oct. 25, 2010), http://www.telegraph.co.uk/technology/social-media/8085772/People-more-likely-to-lie-on-Twitter-than-in-real-life-survey-reveals.html [https://perma.cc/227X-R3SM] (citing Optimum Research survey that found that one-third of the 2012 people surveyed were more honest during face-to-face conversations than on social media); *see also* Bierman v. Weier, 826 N.W.2d 436, 457 (Iowa 2013) (noting that individuals on the internet "have fewer incentives to self-police the truth of what they are saying" because they speak pseudonymously or anonymously and "care less about their reputation for veracity").

61. Nick Bilton, *Disruptions: Twitter's Uneasy Role in Guarding the Truth*, N.Y. TIMES, (Nov. 4, 2012, 11:00 AM), http://bits.blogs.nytimes.com/2012/11/04/disruptions-twitters-faster-gantlet-of-truth/?_php=true&_type=blogs&_r=0 [https://perma.cc/8WTY-L26Q] (quoting David Livingstone Smith, Associate Professor of Philosophy at the University of New England); *cf.* Citizens United v. FEC, 558 U.S. 310, 352 (2010) ("With the advent of the Internet and the decline of print and broadcast media . . . the line between the media and others who wish to comment on political and social issues becomes far more blurred.").

62. Byron Calame, *Scoops, Impact or Glory: What Motivates Reporters?*, N.Y. TIMES, (Dec. 3, 2006), http://www.nytimes.com/2006/12/03/opinion/03pubed.html?pagewanted=all&_r=1& (last visited Dec. 20, 2017); *see also* Bill Grueskin, *In Defense of Scoops*, COLUM. JOURNALISM REV. (Apr. 22, 2013), http://www.cjr.org/united_states_project/in_defense_of_scoops.php [https://perma.cc/82DM-SWXL] ("[B]reaking exclusives can be

"race to the bottom,"[63] causing some journalists to fabricate certain aspects of stories or fail to properly substantiate them. For example, part of the reason why there were so many false news reports during the Boston Marathon bombings is because many reporters were "caught up in the . . . adrenaline of the moment" and motivated by the "thrill [of] being the first . . . to report a story."[64] Being the first to tell a new rumor or gossip story also has tremendous "conversational cash value," which enhances social relationships.[65] These effects are exacerbated for social media users because social media's obsession with speed over content oftentimes leads to impulsive and spontaneous behavior, driven by the fact that our communications are relegated to short bursts of information that effectively rob us of "the richness of human experience and reflection."[66] Thus, social media's "limited temporal existence urges us not to develop or sustain lasting concerns but rather to exist in the temporary and fluid realm of our immediate beliefs, attractions and repulsions" without thinking twice.[67] Journalists and laypersons alike therefore become consumed by the "real immediacy and . . . stimulus response" of social media when reporting on news events.[68] President Obama recognized these effects following the Boston Marathon bombings, noting that, "[i]n this age of instant reporting and tweets and blogs, there's a temptation to latch on to any bit of information, sometimes to jump to conclusions."[69] The very structure and functionality of social media encourages *immediacy* of news reporting over *accuracy*.

Third, social media has tremendous reach across its billions of subscribers, and as a result, posts are propagated effortlessly and frequently. Studies have shown, for example, that any retweeted message will reach an average of 1000 Twitter users, irrespective of how many people fol-

---

contagious: One scoop leads to another, and one ephemeral scoop can lead to a bigger, deeper news break.").

63. Danny Bradbury, *Read all About It*, INFOSECURITY, July–Aug. 2010, at 29, 30 (noting that "most reporters . . . tend to want to be 'first' to tell" a story among reporters and that there is "a race going on about who can break the story first").

64. Gladstone & Garfield, *supra* note 37.

65. Bernard Guerin & Yoshihiko Miyazaki, *Analyzing Rumors, Gossip, and Urban Legends Through Their Conversational Properties*, 56 PSYCHOL. REC. 23, 25–27 (2006) (challenging the notion that people tell rumors, gossip, and urban legends to impart information to the listener or alleviate listener anxiety about the topic); *see also* Bernard Guerin, *Language Use as Social Strategy: A Review and an Analytic Framework for the Social Sciences*, 7 REV. GEN. PSYCHOL. 251, 261 (2003) (noting that "being the first one in a group to be able to tell the others some bit of news" signals superior access to resources and helps maintain social relationships).

66. Mitchell R. Haney, *Social Media, Speed, and Authentic Living*, *in* Social Media and the Value of Truth 44 (Berrin Beasley & Mitchell R. Haney eds.,2013).

67. *Id.* at 44–45.

68. Gladstone & Garfield, *supra* note 37.

69. Barack Obama, President of the United States, Statement by the President (Apr. 19, 2013, 10:05 PM), https://obamawhitehouse.archives.gov/the-press-office/2013/04/19/statement-president [https://perma.cc/4Y9W-NDPK].

lowed the original tweet.[70] The vast reach of social media leads to what Professor Cass Sunstein calls "social cascades,"[71] in the form of a dangerous game of "telephone," whereby lies and rumors are reposted with a click of a mouse.[72] For this reason, the spread of a false report on social media has been analogized to the spread of a virus: "Infected Internet users, who may have picked up bogus info from an inaccurate media report, another person on social media or word-of-mouth, proceed to 'infect' others with each false tweet or Facebook post."[73] Thus, even though social media is useful during emergencies and crises, some regard it as not "reliable, deep or broad enough to meet the information needs of professional organizations, more likely to rely on professional reporters, not unsubstantiated accounts from ordinary citizens."[74]

Some people are likely to believe these lies, at least during times of crises, when fear, anxiety, and uncertainty abound. In such circumstances, people may be susceptible to the false reports of their fellow social media subscribers. As some scholars have noted, social media posts made during moments of crisis that leverage peoples' fears cause users to lose their judgment and "spread facts that are obviously wrong under the pressure of these feelings."[75] Substantiating a post may also be especially difficult when the professional organizations that we rely upon to report accurate news treat the post as accurate without first substantiating it. For this reason, courts and commentators alike have noted that trying to rely on social media account postings "as proof of facts, actually things that have happened, just can't be done."[76]

Importantly, however, rumors and lies propagated via social media are fleeting in time if not in reach. This is because social media acts as a self-correcting (or "crowd-correcting") network. Although social media allows rumors to spread "at great speed," it "has an equal and opposite power to dispel them."[77] Social media communities can oftentimes sub-

---

70. Kwak et al., *supra* note 28.

71. CASS R. SUNSTEIN, REPUBLIC.COM 80 (2002).

72. *See* Hitlin, *supra* note 48, at 3 (noting that false information on the internet can "spread throughout other Web sites and emails" and can "become part of a public folklore even if there are no facts to support the original reports").

73. Victor Luckerson, *Fear, Misinformation, and Social Media Complicate Ebola Fight*, TIME (Oct. 8, 2014), http://time.com/3479254/ebola-social-media/ [http://perma.cc/8TKH-4XCC].

74. Mills et al., *supra* note 35, at 21.

75. *See* Luckerson, *supra* note 73.

76. R.M. v. D.Z., 2013 Ill. App. 3d 120846-U, at *6 (Ill. App. Ct. Mar. 4, 2013); *see also* Matot v. CH, 975 F. Supp. 2d 1191, 1196 (D. Or. 2013) (noting that there are numerous fake social media accounts); Caitlin Dewey, *Lies Are Everywhere on the Internet. But This Free Tool Could Potentially Fight Them*, WASH. POST (May 2, 2014), http://www.washingtonpost.com/blogs/style-blog/wp/2014/05/02/lies-are-everywhere-on-the-internet-but-this-free-tool-could-potentially-fight-them/ [https://perma.cc/6DXS-7M7L] (acknowledging that social media is full of lies).

77. Jonathan Richards & Paul Lewis, *How Twitter Was Used to Spread — and Knock Down — Rumours During the Riots*, GUARDIAN (Dec. 7, 2011),

stantiate a story via a simple Google search — each subscriber has a world of knowledge at her fingertips, which can be used to either verify or discredit any false report in a matter of minutes. Thus, the research costs of substantiating a particular post are relatively low in cyberspace.[78] No longer will a lie "travel halfway around the world before the truth puts its shoes on," because "lies get slapped down really fast" in the social media world.[79]

This crowd-correcting mechanism was evident following the July 2016 Dallas shootings in which a man ambushed and fired upon a group of police officers, killing five and injuring nine others. In the wake of the shootings, the Dallas Police Department tweeted the photograph of Mark Hughes, the man they believed was the perpetrator. But "within minutes," people began tweeting evidence, including video showing Hughes on the street with the crowd after shots were fired, proving that he was not in fact the gunman.[80] Thus, while lies spread on social media

---

http://www.theguardian.com/uk/2011/dec/07/how-twitter-spread-rumours-riots [https://perma.cc/E8V6-XE9Q] (citing a *Guardian* & London School of Economics study of the 2011 London riots).

78. *See* Michael R. Baye et al., *The Evolution of Product Search*, 9 J.L. ECON. & POL'Y 201, 204 (2013) (noting that "the internet arguably has reduced product search costs more than any innovation in the history of mankind"); Alex Aferiat, Note, *It's Google's World and We're Just Clicking in It: Why the Growth of Sponsored Link Advertising Necessitates a Shift of Trademark Regulation on the Internet*, 47 NEW ENG. L. REV. 157, 175 (2012) (noting that "consumers expend virtually no extra 'search costs'" because searching for the desired information "on Google or another search engine is both expeditious and effortless"). *But see* Bierman v. Weier, 826 N.W.2d 436, 454 (Iowa 2013) ("[Defendants] argue that the Internet is 'a great equalizer' and has rendered libel per se obsolete because the targets of defamation can respond quickly at minimal cost. We are not persuaded, however, that the Internet's ability to restore reputations matches its ability to destroy them.").

79. Gerry Shih, *During Hurricane Sandy, Twitter Proves a Lifeline Despite Pranksters Like @ComfortablySmug*, HUFFINGTON POST (Oct. 31, 2012), http://www.huffingtonpost.com/2012/10/31/hurricane-sandy-twitter-comfortablysmug_n_2047754.html [https://perma.cc/7Z8H-5UY3] (quoting Ben Smith, editor at Buzzfeed, Inc.); *see also* DAVIS III ET AL., *supra* note 30 (summarizing how the Boston Police Department refuted rumors after the Marathon bombing). One study, however, challenges the crowd-correcting theory of social media, finding some "evidence of crowd-correction for each rumor but with considerably smaller proportions of correction" than previously believed. Kate Starbird et al., *Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter After the 2013 Boston Marathon Bombing*, iConference 2014 Proc. at 661 (2014), https://www.ideals.illinois.edu/bitstream/handle/2142/47257/308_ready.pdf [https://perma.cc/3YPR-MX8U].

80. *See* Lisa Gutierrez, *How Twitter Went to Bat for Mark Hughes, Misidentified as a Suspect in Dallas Police Shootings*, KAN. CITY STAR (July 8, 2016), http://www.kansascity.com/news/nation-world/national/article88409497.html [https://perma.cc/23KF-V7LG]. The Hughes case demonstrates the crowd-correcting mechanism operating at its best. But some research shows that this crowd-correcting mechanism often lags significantly behind the misinformation, sometimes by as much as twenty hours. *See* Chengcheng Shao & Giovanni Luca Ciampaglia, *Hoaxy: A Platform for Tracking Online Misinformation*, PROC. OF 25TH INT'L CONF. COMPANION ON WORLD WIDE WEB (2016); *see also* Michela Del Vicario et al., *The Spreading of Misinformation Online*, 113 PROC. NAT'L ACAD. SCIS. 554 (Jan. 19, 2016) (showing that 25% of lies posted on social media last for ten or more hours).

certainly have potential to cause harm, social media has an inherent countermeasure that may mitigate the scope of the harm at least in some circumstances.

## III. CRIMINALIZING FALSE SPEECH ON SOCIAL MEDIA: FALSE REPORTING STATUTES

### *A. Repercussions for False Reports on Social Media*

While much scholarly work attempts to determine how to identify false news-related tweets,[81] little work has been done to identify and analyze legal implications for spreading such reports.[82] Can the author of a false tweet be held *criminally* liable?[83] This question is not merely academic, as false internet reports can have very real and harmful effects. For example, in response to the rumor that President Obama had been injured in explosions at the White House, the Dow Jones Index plunged over 140 points, and the S&P 500 Index declined 0.9%, which is "enough to wipe out $130 billion in stock value in a matter of seconds."[84] And in the wake of the Boston Marathon bombings, social media accounts cropped up seeking to turn a quick profit by claiming to raise money for victims through fraudulent charity funds.[85] Such fake post-disaster, cyber-based charities are not new. The New Jersey Attor-

81. *See, e.g.*, *supra* notes 35, 48. In fact, methods and systems for detecting lies on social media are now being patented. *See, e.g.*, U.S. Patent Nos. 9,361,382 (filed May 30, 2014), 9,047,253 (filed Mar. 13, 2013).

82. *But see generally* John R. Grasso, *Criminal Consequences of Sending False Information on Social Media*, 60 R.I. B.J. 5 (2011) (discussing viability of First Amendment challenge to Rhode Island's computer crimes law); Daniel S. Harawa, *Social Media Thoughtcrimes*, 35 PACE L. REV. 366 (2014) (analyzing distinction between protected and criminal expression made on social media); Erin D. Guyton, Comment, *Tweeting "Fire" in a Crowded Theater: Distinguishing Between Advocacy and Incitement in the Social Media World*, 82 MISS. L.J. 689 (2013) (analyzing First Amendment restrictions for advocacy and incitement speech made on social media).

83. A social media user can, in certain circumstances, be civilly liable for libelous and defamatory posts. *See* Ellyn M. Angelotti, *Twibel Law: What Defamation and its Remedies Look Like in the Age of Twitter*, 13 J. HIGH TECH. L. 430 (2013). This Article does not suggest, however, that content providers should be liable for false posts made by their subscribers. In fact, the Communications Decency Act would likely bar any attempt to hold providers civilly liable (under any law) or criminally liable (except under a federal criminal statute) for false content originated by their users. *See* Communications Decency Act § 230(c), 47 U.S.C. § 230(c) (2012).

84. Christopher Matthews, *How Does One Fake Tweet Cause a Stock Market Crash*, TIME (Apr. 24, 2013), http://business.time.com/2013/04/24/how-does-one-fake-tweet-cause-a-stock-market-crash/ [https://perma.cc/3Y5D-9LCX]; Heidi Moore & Dan Roberts, *AP Twitter Hack Causes Panic on Wall Street and Sends Dow Plunging*, GUARDIAN (Apr. 21, 2013), http://www.theguardian.com/business/2013/apr/23/ap-tweet-hack-wall-street-freefall [https://perma.cc/3AQJ-YE5U].

85. Melanie Hicken, *Beware Bogus Boston Marathon Charity Websites*, CNN MONEY (Apr. 17, 2013), http://money.cnn.com/2013/04/17/pf/boston-marathon-charity/ [https://perma.cc/E3KG-EY62].

ney General and Division of Consumer Affairs initiated legal proceedings to shut down a website for the Hurricane Sandy Relief Foundation, which raised more than $630,000 in cash donations but gave less than 1% to victims of the disaster.[86]

Some countries have demonstrated a willingness to prosecute the authors of false internet reports. In Mexico, two "Twitter terrorists" were criminally prosecuted and faced thirty years in prison for spreading rumors about fake school shootings.[87] And in England, two people were sentenced to four years in prison for spreading false information through posts on Facebook during the 2011 riots,[88] while another was sentenced to twelve weeks in jail for posting offensive comments on Facebook about a missing five-year-old girl.[89] Further, in the high-profile "Twitter Joke Trial," an accountant was convicted of sending a menacing tweet for his tongue-in-cheek joke about "blowing [an] airport sky high," though he eventually succeeded in having his conviction reversed in a closely-watched appeal to the High Court of Justice.[90] In the United Kingdom alone, 653 people were charged for "social networking crimes" in 2011.[91]

These convictions raise serious questions about freedom of speech on social media. Commentators have noted that the outcomes in some of these cases would have been different in the United States, where free speech rights are broader and enjoy strong constitutional protection.[92] But while criminal prosecutions for social media activity are infrequent

---

86. *Id.*

87. Jo Adetunji, *"Twitter Terrorists" Face 30 Years After Being Charged in Mexico*, GUARDIAN (Sept. 4, 2011), http://www.theguardian.com/world/2011/sep/04/twitter-terrorists-face-30-years [https://perma.cc/RX9R-PBGL].

88. Owen Bowcott et al., *Facebook Riot Calls Earn Men Four-Year Jail Terms Amid Sentencing Outcry*, GUARDIAN (Aug. 16, 2011), http://www.theguardian.com/uk/2011/aug/16/facebook-riot-calls-men-jailed [https://perma.cc/7NF2-3BTJ].

89. Press Ass'n, *April Jones Murder: Teenager Jailed Over Offensive Facebook Posts*, GUARDIAN (Oct. 8, 2012) https://www.theguardian.com/uk/2012/oct/08/april-jones-teenager-jailed-facebook [https://perma.cc/6R5N-NU24].

90. Owen Bowcott, *Twitter Joke Trial: Paul Chambers Wins High Court Appeal Against Conviction*, GUARDIAN (July 27, 2012), http://www.theguardian.com/law/2012/jul/27/twitter-joke-trial-high-court [https://perma.cc/TH2U-4TAE]; *see also* Salma Abdelaziz, *Teen Arrested for Tweeting Airline Terror Threat*, CNN (Apr. 14, 2014), http://www.cnn.com/2014/04/14/travel/dutch-teen-arrest-american-airlines-terror-threat-tweet/ [https://perma.cc/8A6R-JDAD] (discussing the arrest and charges filed under Dutch law against a fourteen-year-old Dutch girl for posting a false or alarming announcement after she jokingly tweeted a terror threat to American Airlines).

91. Becky Evans, *5,000 People Investigated by Police for Something They Said on Facebook or Twitter as "Social Network Crime" Soars 800%*, DAILY MAIL ONLINE (Dec. 27, 2012), http://www.dailymail.co.uk/news/article-2253692/Facebook-Twitter-crime-sees-fold-increase-police-deal-5-000-cases-involving-websites.html (last visited Oct. 23, 2017).

92. *See* Jeff John Roberts, *Repeat a Horrible Lie on Twitter, Pay $25,000: Is That Fair?*, GIGAOM (Oct. 26, 2013), http://gigaom.com/2013/10/26/repeat-a-horrible-lie-on-twitter-pay-25000-is-that-fair/ [https://perma.cc/5MHS-4YXX] (discussing case in which a U.K. man was fined $25,000 after he retweeted a claim that wrongly identified a British lord as a child molester, and noting that, "as yet, no one in America has tried to sue over a retweet").

in the United States, they are not nonexistent. In 2012, for example, an Ohio teenager was convicted of "inducing panic" for his Facebook posts stating in the immediate aftermath of the Sandy Hook shooting that "there needs to be another mass murder."[93] Ultimately, the Ohio Court of Appeals affirmed his conviction, finding that

> [The] teen's Facebook posts caused members of the public to contact police, required weekend meetings between the police, Principal Carey, Wilmington school district's superintendent and the school district's business manager, led to the school issuing an 'all call,' alerting the entire student body to the situation, triggered a police presence at Wilmington High School on the following day of classes, and resulted in several students being absent from school due to their parents' fear of what might happen. These responses to [the teen's] Facebook posts are sufficient to show serious public inconvenience and alarm.[94]

While the teenager's Facebook posts went well beyond mere falsehoods and involved actual threats of violence, the Ohio case nonetheless provides an example in which a court recognized that speech made on social media can have real-life consequences.

States have only recently begun prosecuting harmful social media activity, but statutes for addressing false speech in public channels have been on the books in many states for decades. These statutes — sometimes called "false reporting statutes" — proscribe the circulation of false reports of criminal activity or natural catastrophe or disaster to the public, but they are seldom used in the cyber context.[95] That will likely change, however, as social media becomes even more established as a dominant source of news. In fact, currently before the New York Assembly is a bill that proposes to increase the severity of the offenses in the New York false reporting statute analyzed in this article, spurred by concern about the unique harms inflicted through online communication.[96]

---

93. *In re* P.T., 995 N.E.2d 279, 281 (Ohio Ct. App. 2013).

94. *Id.* at 286.

95. In addition, many states have cyberbullying, cyberstalking, terroristic threat, false statement, and hoax statutes. *See* Ira P. Robbins, *Anthrax Hoaxes*, 54 AM. U. L. REV. 1 (2004) (summarizing states' false reporting statutes and comparing them to hoax and terroristic threat statutes); *see also* Kim Zetter, *Judge Acquits Lori Drew in Cyberbullying Case, Overrules Jury*, WIRED (July 2, 2009), http://www.wired.com/2009/07/drew_court/ [https://perma.cc/P8B7-BRM6] (discussing reversal of 2008 conviction of woman who created a fake MySpace profile to cyberbully a thirteen-year-old girl who later committed suicide).

96. 2017 New York Assembly Bill No. 8749 §§ 1, 7 ("Technological innovations have resulted in various platforms for personal sharing, many of which are often misused maliciously.

### *B. False Reporting Statutes' Derivation and Theoretical Underpinnings*

Most states' false reporting statutes derive in part from the American Law Institute's Model Penal Code, which adopted a criminal provision for "false public alarm" in 1962. That provision stated:

A person is guilty of a misdemeanor if he initiates or circulates a report or warning of an impending bombing or other crime or catastrophe, knowing that the report or warning is false or baseless and that it is likely to cause evacuation of a building, place of assembly, or facility of public transport, or to cause public inconvenience or alarm.[97]

The provision updated and codified older offenses against public order — i.e., those that "affect a large number of defendants, involve a great proportion of public activity, and powerfully influence the view of public justice held by millions of people."[98] Those offenses include false fire alarms, false reports of crime, and false warnings of bomb plantings and similar incidents.[99]

Notably, the provision's *mens rea* requirement limits the reach of the statute in two important ways. First, the provision imposes an intent requirement with respect to the veracity of the report. Specifically, the provision "requires that the actor initiate or circulate a report or warning known by him to be false. Thus, the provision does not reach the individual who merely repeats a rumor or otherwise circulates information that he does not know to be baseless."[100] Second, the provision imposes an intent requirement with respect to the ensuing harm: "The actor must . . . know that his conduct is 'likely to cause evacuation of a building, place of assembly, or facility of public transport, or to cause public inconvenience or alarm.'"[101] Thus, excluded from liability is "the practical joker or other person who circulates a false alarm in circumstances where he is unaware of the potential for serious consequences."[102] As discussed *infra*, New York's false reporting statute does not impose a *mens rea* requirement with respect to the ensuing harm.[103]

However, the Model Penal Code provision is also *broader* than some false reporting statutes because it does not impose a requirement that the false report be made to a particular audience — e.g., a government official. This is because the Model Penal Code provision "is to guard against the inconvenience and alarm that may be occasioned by

---

Current penal laws are centered on harm that occurs within a public setting, which fails to account for the expansive and dynamic nature of modern technology.").

97. MODEL PENAL CODE § 250.3 (AM. LAW INST. 1980).

98. MODEL PENAL CODE AND COMMENTARIES, PT. II, at 309 (AM. LAW INST. 1980).

99. *Id.* at 355.

100. *Id.* at 356.

101. *Id.*

102. *Id.*

103. *See* N.Y. PENAL LAW § 240.50(1) (McKinney 2016).

circulating a false alarm directly to members of the public," generally.[104] Thus, as the commentaries accompanying the Model Penal Code make clear, the statute would apply to "Mr. Justice Holmes' famous example of the person who cries 'fire' in a crowded theater."[105]

As of 1980, seven states had enacted laws substantially identical to the Model Penal Code offense, while four others had proposed such provisions.[106] Today, most states have false reporting statutes, many of which are similar to the Model Penal Code. But some states' statutes are significantly broader. For example, Delaware's and Kentucky's false reporting statutes impose liability for circulating a knowingly false report that is likely to cause public alarm or inconvenience,[107] whether the speaker knows about the likelihood of harm or not.

### *C. New York's False Reporting Statute: A Blunt Tool for Combating False Speech*

New York's false reporting statute is perhaps the broadest in the United States. New York Penal Law § 240.50 addresses "falsely reporting an incident in the third degree," and states:

> A person is guilty of falsely reporting an incident in the third degree when, knowing the information reported, conveyed or circulated to be false or baseless, he or she[] . . . [i]nitiates or circulates a false report or warning of an alleged occurrence or impending occurrence of a crime, catastrophe or emergency under circumstances in which it is not unlikely that public alarm or inconvenience will result[.][108]

Falsely reporting an incident in the third degree is a class A misdemeanor and is punishable by up to one year's imprisonment and a $1,000 fine.[109] New York law also allows entities providing emergency services to seek restitution for "the amount of funds reasonably expended for the purpose of responding" to false reports.[110]

Section 240.50(1), enacted in 1965, was designed to augment offenses that proscribed giving false fire alarms and circulating false

---

104. MODEL PENAL CODE AND COMMENTARIES, PT. II, at 356 (AM. LAW INST. 1980).
105. *Id.* at 357.
106. *Id.* at 357–58.
107. 11 DEL. CODE § 1245(1); KY. PENAL CODE § 519.040(1)(e).
108. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).
109. *Id.* §§ 70.15(1), 80.05(1). See also 2017 New York Assembly Bill No. 8749 § 7, which proposes to increase the classification of falsely reporting an incident in the third degree from a class A misdemeanor to a class E felony, punishable by up to four years' imprisonment and a $5,000 fine. N.Y. PENAL LAW §§ 70.00, 80.00.
110. *Id.* § 60.27(13).

"bomb scare" reports.[111] As enacted, however, Section 240.50(1) encompasses more than just those offenses because it includes false reports or warnings concerning any "crime, catastrophe or emergency." Notably, like the Model Penal Code, Section 240.50(1) is broad in three additional respects. First, it proscribes false reports that could cause a mere "public inconvenience" rather than a more serious degree of harm. Second, it does not require the report to be made to any particular person or agency. Third, it does not require *actual* public alarm or inconvenience, but instead requires only that such alarm or inconvenience be "not unlikely" to result.[112]

In fact, Section 240.50(1) is *broader* than the Model Penal Code in some ways because, although the statute requires knowledge that the statement is false, it does not require knowledge or intent with respect to the ensuing public alarm or inconvenience. Further, in contrast to the Model Penal Code and some other states' laws, which require that the false report be "likely to cause" harm,[113] Section 240.50(1) requires that the false report merely be "*not un*likely" to cause harm. The former provides a reasonable nexus between the *actus reus* and the ensuing harm. The latter provides only a tenuous nexus, as it encompasses false reports that *could conceivably* cause public alarm or inconvenience but that do not "likely" cause such harm.

Yet, despite the breadth of the statute, few First Amendment challenges have been lodged. Only one reported case has addressed a First Amendment challenge and rejected it. In *People v. Hanifin*,[114] the Supreme Court of New York, Appellate Division, upheld a conviction of a man who parked his car in the middle of Main Street in Union, New York, climbed on top of his vehicle, doused himself with what appeared to be gasoline (but was actually water), and called 911, threatening to set himself on fire if the war in Iraq did not end by a certain time that day.[115] The defendant argued that he was "conducting a protest" under the First Amendment, but the court rejected that argument.[116] Citing Justice Holmes's admonition against shouting "fire!" in a crowded theater,[117] the court perfunctorily concluded that the defendant's First

111. *Id.* § 240.50(1) (Practice Commentary).

112. If, however, an emergency response member suffers serious physical injury or death while responding to the false report, the person making the report is guilty of falsely reporting an incident in the first degree. *See id.* § 240.60. Falsely reporting an incident in the first degree is a class E felony punishable by up to four years' imprisonment and a $5,000 fine. *Id.* §§ 70.00(2)(e), 80.00(1).

113. MODEL PENAL CODE § 250.3; *see also, e.g.*, 11 DEL. CODE § 1245(1); KY. PENAL CODE § 519.040(1)(e); OHIO REV. CODE § 2917.32(A)(1); W. VA. CODE § 61-6-20(1).

114. 910 N.Y.S.2d 212 (2010).

115. *Id.* at 213.

116. *Id.* at 214.

117. *See* Schenck v. United States, 249 U.S. 47, 52 (1919).

Amendment rights "do not permit him to falsely report an impending fire."[118]

No court has engaged in a robust First Amendment analysis of Section 240.50(1), let alone in response to a challenge involving speech conducted on social media. But such a challenge may be imminent. Indeed, in 2014, a teenager was convicted of violating Oregon's disorderly conduct statute for engaging in a conversation on MySpace about shooting up a local high school.[119] Like New York's false reporting statute, the Oregon statute imposes liability on reports known to be false concerning "an alleged or impending fire, explosion, catastrophe or other emergency."[120] But unlike New York's statute, the Oregon statute contains a *mens rea* element with respect to the ensuing harm, imposing liability only if the speaker *intends* "to cause public inconvenience, annoyance or alarm, or knowingly creat[es] a risk thereof . . . ."[121] The trial court rejected the teenager's First Amendment defense, but the Oregon Court of Appeals never reached the constitutional question, deciding instead that the teenager could not be liable because he merely responded to another's post and therefore did not knowingly initiate and circulate the report.[122] The court also determined that there was "no evidence to support the inference that defendant knew that his contribution to the conversation would ultimately move beyond the conversation itself so as to cause the specified risks."[123] The court reversed the conviction on these bases.

Such cases suggest that it is likely only a matter of time before a court squarely addresses the constitutionality of false reporting statutes as applied to false speech communicated via social media.

## IV. THE FIRST AMENDMENT'S ROLE IN REGULATING FALSE SPEECH

This Part analyzes the viability of a First Amendment challenge to Section 240.50(1) as applied to false speech on social media, beginning with a description of First Amendment doctrine generally, and its role in regulating false speech.

118. *Hanifin*, 910 N.Y.S.2d at 214.
119. State v. Nelson, 341 P.3d 787 (Or. Ct. App. 2014).
120. OR. REV. STAT. § 166.023(1)(a) (2016).
121. *Id.* at 790.
122. *Nelson*, 341 P.3d at 788, 790.
123. *Id.*

*A. Theoretical Underpinnings: Testing Truth in the Marketplace*

The First Amendment to the Constitution states that "Congress shall make no law . . . abridging the freedom of speech."[124] While American jurisprudence has rejected an absolutist interpretation of the Amendment, freedom of speech remains "a preeminent constitutional value supported by multiple justifications,"[125] the most resonant being the marketplace of ideas. Articulated by Justice Oliver Wendell Holmes in his dissenting opinion in *Abrams v. United States*,[126] this theory explains freedom of speech in terms of an open marketplace in which ideas compete against one another for acceptance by the public. "[T]he best test of *truth*," Holmes wrote, "is the power of the thought to get itself accepted in the competition of the market."[127] The theory has been absorbed into the legal culture, and Justices' iterations of the idea permeate First Amendment jurisprudence.

Under the marketplace of ideas model, a commitment to democratic government and individual liberty requires that repugnant, false, or otherwise misleading speech be allowed to compete unrestrained with other speech. In *Cohen v. California*, for example, the Court confirmed that the marketplace of ideas is central to a free society, as it overturned the conviction of a defendant who had worn a jacket bearing the words "Fuck the Draft" in a courthouse in violation of California's breach of the peace statute.[128] In particular, the Court noted that:

> [T]he First Amendment is designed and intended to remove governmental restraints from the arena of public discussion, putting the decision as to what views shall be voiced largely *into the hands of each of us* . . . in the belief that no other approach would comport with the premise of individual dignity and choice upon which our political system rests.[129]

The Court also rejected the assertion that the state could censor to cleanse public discourse: "That the air may at times seem filled with verbal cacophony is, in this sense not a sign of weakness but of strength," Justice Harlan wrote for the Court.[130] He continued: "We cannot lose sight of the fact that, in what otherwise might seem a trifling

124. U.S. CONST. amend. I.
125. RODNEY A. SMOLLA, SMOLLA AND NIMMER ON FREEDOM OF SPEECH § 2:8 (2017).
126. 250 U.S. 616 (1919).
127. *Id.* at 630 (Holmes, J., dissenting) (emphasis added).
128. 403 U.S. 15, 26 (1971).
129. *Id*. at 24 (emphasis added).
130. *Cohen*, 403 U.S. at 25. *But see* FCC v. Pacifica Found., 438 U.S. 726 (1978) (holding that the FCC could restrict indecent language over broadcast radio).

and annoying instance of individual distasteful abuse of a privilege, these fundamental societal values are truly implicated."[131] As such, "[t]he marketplace theory is thus best understood *not* as a guarantor of the final conquest of truth, but rather as a defense of the *process* of an open marketplace of speech," where false speech can be tested and refuted.[132] John Stuart Mill referred to this ability of the marketplace to refute falsehoods as a "collision with error," which he noted leads to a "clearer perception and livelier impression of truth."[133]

The marketplace model is particularly well suited for application to speech on social media. As discussed above, social media increasingly facilitates the process of open debate. As a fluid and easily accessible forum that encourages immediacy and acts as a self-correcting network, social media platforms literally put the decision as to what shall be voiced in the hands of each of us.[134] Technological advancements do not alter the basic values of the First Amendment, but expand the marketplace of ideas to new frontiers. Indeed, this concept was recognized by the Supreme Court in *Reno v. ACLU*,[135] where the Court struck down a statute that criminalized the communication of obscene, patently offensive, or indecent material to minors over the internet.[136] In distinguishing the Communications Decency Act from previously upheld statutes that prohibited indecent speech,[137] the Court agreed with the notion that "the content on the Internet is as diverse as human thought" and found "no basis for qualifying the level of First Amendment scrutiny that should be applied to this medium."[138]

### *B. First Amendment Framework*

Despite the First Amendment's unqualified words, "it is well understood that the right to free speech is not absolute at all times and under all circumstances."[139] Explaining that "[e]ach medium of expression must be assessed for First Amendment purposes by standards suited for it,"[140] the Supreme Court has devised an array of doctrines to analyze

131. *Cohen*, 403 U.S. at 25.
132. SMOLLA, *supra* note 125, at § 2:19.
133. JOHN STUART MILL, ON LIBERTY 36 (1858).
134. *Cf.* Abrams v. United States, 250 U.S. 616, 630 (1919).
135. 521 U.S. 844 (1996).
136. *Id.* at 885.
137. *See, e.g.*, *FCC v. Pacifica Found.*, 438 U.S. at 726.
138. *Reno*, 521 U.S. at 870. In striking down the CDA as overbroad, the Court took particular issue with the term "indecent" and the fact that the statute imposed a criminal penalty for speech.
139. Chaplinsky v. New Hampshire, 315 U.S. 568, 571 (1942).
140. Se. Promotions v. Conrad, 420 U.S. 546, 557 (1975).

federal and state[141] governmental action abridging many areas of speech. The constitutional inquiry requires a court to determine whether the law (1) regulates a category of speech that is unprotected under the First Amendment or enjoys something less than full protection, giving the government the regulatory authority, and whether the law (2) is a content-based restriction — which are presumed invalid under strict scrutiny — or a content-neutral restriction — which are subject to intermediate scrutiny, a less speech-protective test.

1. The First Amendment Does Not Protect Certain Categories of Low-Value Speech

With regard to the first inquiry, some categories of speech are typically treated as lying outside of full First Amendment protection. On the authority of English common law, the Court has determined that "[t]here are certain well defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem."[142]

Indeed, the Supreme Court has upheld restrictions on speech in several historic categories including incitement, libel, obscenity, defamation, speech integral to criminal conduct, fighting words, child pornography, fraud, true threats, and speech presenting some grave and imminent threat the government has the power to prevent.[143] If the regulated speech falls within a circumscribed category, the Court most often submits the regulation to rational basis review, a highly deferential standard under which a law is almost always upheld.[144]

2. Content-Based Restrictions Are Subject to Heightened Scrutiny

Assuming that the speech at issue does not fall within an unprotected category, the second inquiry asks whether the speech is "content-based" or "content-neutral." Although the distinction is sometimes difficult for courts to make,[145] the "principal inquiry" in determining whether a law is content-based or content-neutral is "whether the government has adopted a regulation of speech because of [agreement or] disagreement

141. *See* Gitlow v. People of New York, 268 U.S. 652, 666 (1925) (extending constitutional protection of freedom of speech to the states through the incorporation of the First Amendment under the Due Process Clause of the Fourteenth Amendment).

142. *Chaplinsky*, 315 U.S. at 571–72.

143. *See, e.g.*, United States v. Alvarez, 567 U.S. 709, 717 (2012).

144. *See* JOHN NOWAK & RONALD ROTUNDA, CONSTITUTIONAL LAW 574–75 (4th ed. 1991) ("[T]he Court will ask only whether it is conceivable that the classification bears a rational relationship to an end of government which is not prohibited by the Constitution. So long as it is arguable that the other branch of government had such a basis for creating the classification a court should not invalidate the law.").

145. *See* SMOLLA *supra* note 125, at § 3:1.

with the message it conveys."[146] "[A]bove all else, the First Amendment means that the government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[147] Content-based laws do just that. Therefore, content-based restrictions on protected speech outside of the historically unprotected categories discussed above are presumed invalid, and the government bears the burden of proving their constitutionality.[148] Subject to "strict scrutiny," the law will be tolerated only upon a showing that it is narrowly tailored to promote a compelling government interest.[149]

A rare law to have passed the test was a Tennessee provision that prohibited the solicitation of votes and the display or distribution of campaign materials within 100 feet of the entrance to a polling place. In *Burson v. Freeman*, a 5-3 Court held that the campaign-free zones — content-based restrictions on political speech — served the state's compelling interest in protecting citizens' "right to vote freely and effectively,"[150] and since the prescribed area was not so large as to completely block out political messages, the statute was sufficiently tailored.[151] Most content-based restrictions, however, do not survive under this speech-protective standard.[152]

A regulation of unprotected speech may still violate the First Amendment's rule against content discrimination if it draws distinctions among *subcategories* of speech that cannot be justified. In *R.A.V. v. City of St. Paul*,[153] for example, the Court invalidated a restriction governing certain "fighting words," an area the Court recognizes as low-value speech.[154] The law at issue prohibited the display of a symbol which one knows or has reason to know "arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender."[155] A unanimous Court held that the ordinance was facially invalid[156] content dis-

146. Ward v. Rock Against Racism, 491 U.S. 781, 791 (1989).

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

148. The Constitution "demands that content-based restrictions on speech be presumed invalid . . . and that the Government bear the burden of showing their constitutionality." Ashcroft v. ACLU, 542 U.S. 656, 660 (2004).

149. *See, e.g.*, Burson v. Freeman, 504 U.S. 191, 199 (1992).

150. 504 U.S. at 208.

151. *See id*. at 208–10.

152. *See, e.g.*, *id.* at 211 ("In conclusion, we reaffirm that it is the rare case in which we have held that a law survives strict scrutiny.").

153. 505 U.S. 377 (1992).

154. *Id.* at 381; *see also* Chaplinsky v. New Hampshire, 315 U.S. 568*,* 572 (1942) (explaining that fighting words are those that, by their very utterance, "inflict injury or tend to incite an immediate breach of the peace").

155. *R.A.V.*, 505 U.S. at 380.

156. Generally, a plaintiff can only succeed in a facial challenge by establishing that no set of circumstances exists under which the Act would be valid. *See* SMOLLA, *supra* note 125, at § 16:25.50. But "[i]n the First Amendment context, [c]riminal statutes . . . those that make unlawful a substantial amount of constitutionally protected conduct may be held facially invalid even if they also have legitimate application." The court stated that "The St. Paul antibias ordi-

crimination under the First Amendment. By limiting specific classes of fighting words — those based on "race, color, creed, religion or gender" — the government had impermissibly expressed a "special hostility towards the *particular* biases thus singled out."[157] The law was not narrowly tailored to serve the compelling governmental interest in protecting the community against "bias-motivated threats to public safety and order," since an ordinance not limited to those classes would have had the same beneficial effect.[158] The Court analogized the regulation to that of another unprotected category of speech: "[T]he government may proscribe libel; 'but it may not . . . [proscribe] only libel critical of the government.'" [159] In short, unprotected speech categories cannot be made "the vehicles for content discrimination unrelated to their distinctively proscribable content."[160]

Laws that confer benefits or impose burdens on speech without reference to the ideas or views expressed receive greater protection. These content-neutral restrictions still have the effect of reducing the total quantity of speech in the market, but they do not pose the same inherent dangers to free expression as content-based regulations; thus they are subject to a less rigorous analysis. A content-neutral regulation will usually be sustained if it withstands the First Amendment "intermediate scrutiny" standard set forth in *United States v. O'Brien*[161] — i.e., if it advances important governmental interests unrelated to suppression of free speech and does not burden substantially more speech than necessary to further those interests.[162] Primary examples of such laws include regulation of (1) activities that have a non-speech component (e.g., an executive agency rule that requires cable operators to carry the signals of local broadcasters),[163] (2) secondary effects (e.g., zoning laws that restrict the location of adult entertainment enterprises),[164] and (3) the time, place, or manner of speech in a public forum (e.g., policies that prohibit public speaking in a public park or on a highway).[165] When reviewing regulation purporting to regulate the last two categories, the Court also requires that alternate communication channels remain open.

---

nance is such a law. Although the ordinance reaches conduct that is unprotected, it also makes criminal expressive conduct that causes only hurt feelings, offense, or resentment, and is protected by the First Amendment. The ordinance is therefore fatally overbroad and invalid on its face." *R.A.V.,* 505 U.S. at 414 (citations omitted).

157. *Id.* at 396 (emphasis added).

158. *Id.* at 395–96.

159. *Id.* at 384.

160. *Id.* at 383–84.

161. 391 U.S. 367 (1968).

162. *See id*. at 377.

163. Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622 (1994).

164. City of Renton v. Playtime Theaters, Inc., 475 U.S. 41 (1986).

165. Davis v. Massachusetts, 167 U.S. 43 (1897); RUSSELL L. WEAVER & DONALD E. LIVELY, UNDERSTANDING THE FIRST AMENDMENT 109 (2d ed. 2006).

### *C. The First Amendment Protects Some Types of Harmful Speech*

Though modern First Amendment jurisprudence sometimes permits speech to be penalized when it causes harm, not all injuries qualify as harms sufficient to justify regulation of speech. Over time, the Court has raised the bar for what qualifies as a speech-suppression rationale for the category of speech that likely has the potential to do the most harm — incitement to violence or lawless action. In *Schenck v. United States*, a World War I-era case, the Court upheld the defendant's conviction under the Espionage Act of 1917 for causing insubordination of military forces by circulating a pamphlet to draftees telling them to obstruct the draft.[166] In so doing, Justice Holmes announced the "clear and present danger" test: "The question in every case," he declared, "is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent."[167] Justice Holmes's classic line maintained that "[t]he most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[168]

While his example endures, the clear and present danger test amounts to a mere intent and bad tendency test in practice.[169] The Court now requires a critical assessment of the practical consequences of the regulated speech. Speech advocating the use of force or crime can only be proscribed where (1) the speech is "directed to inciting or producing imminent lawless action" — a requirement of intent; and (2) the advocacy is also "likely to incite or produce such action."[170] Importantly, when the Court examines the strength of the government interest proffered today, it "unmistakably insists that any limit on speech be grounded in a realistic, factual assessment of harm."[171]

One interesting question in light of this First Amendment jurisprudence is how courts ought to treat speech made in jest that has the practical effect of inciting violence or causing harm although it was not intended to incite violence or cause harm. Such speech may serve important social or political functions but seems to fall within Justice Holmes's "fire!" hypothetical.[172] Many examples of false speech on social media fall into this category. For example, in the "Twitter Joke Tri-

166. 249 U.S. 47 (1919).

167. *Id*. at 52.

168. *Id.*

169. *See* United States v. Williams, 553 U.S. 285, 321–22 (2008) (Souter, J., dissenting).

170. Brandenburg v. Ohio, 395 U.S. 444, 447 (1969); *see also* NAACP v. Claiborne Hardware Co., 458 U.S. 886, 927 (1982) ("[M]ere advocacy of the use of force or violence does not remove speech from the protection of the First Amendment.").

171. *Williams*, 553 U.S. at 322 (Souter, J., dissenting).

172. For a discussion of First Amendment restrictions involving advocacy and incitement speech made on social media, see Guyton, *supra* note 82.

al," discussed in Section III.A, *supra*, a man was convicted of sending a menacing tweet for his tongue-in-cheek joke about "blowing [an] airport sky high."[173]

These occurrences are not new, or even unique to social media. Take, for example, Orson Welles's 1938 radio broadcast adaptation of H.G. Wells's *War of the Worlds*, in which Welles reported that Martians had invaded New Jersey. The broadcast caused a "wave of mass hysteria," as thousands of people evacuated their homes and "called the police, newspapers and radio stations here and in other cities of the United States and Canada seeking advice on protective measures against the raids."[174] That type of speech is probably more likely to be protected under the First Amendment. In fact, one district court recently noted that the *War of the Worlds*-style broadcast "stands in a difficult place in First Amendment jurisprudence" because, although such speech "runs the risk of creating considerable public nuisance and unease as described in *Schenk* [sic], . . . it would be difficult to exclude the original *War of the Worlds* broadcast, and the sensational reaction to it, from our modern idea of the marketplace of ideas."[175] Thus, even false speech that has the possibility (or even probability) of causing public harm can have value in some circumstances.

### *D. The First Amendment Protects Some Types of Lies*

While some lies that may incite lawlessness fall outside the First Amendment's reach, not all lies are unprotected. The Supreme Court's most recent ruling on the status of false statements under the First Amendment, *United States v. Alvarez*,[176] held that lies generally qualify as protected speech.[177] In a 6-3 decision, the Court invalidated Section 704(b) of the Stolen Valor Act of 2005,[178] which made it a crime to lie about receiving military medals or honors.[179] The defendant had been convicted of violating the law after he falsely claimed at a public board meeting to have been awarded the Congressional Medal of Honor.[180]

The plurality opinion by Justice Kennedy, joined by Chief Justice Roberts, Justice Ginsburg, and Justice Sotomayor, held that not all proscriptions of false statements are automatically exempt from rigorous First Amendment scrutiny.[181] While content-based restrictions on speech

---

173. *See* Bowcott, *supra* note 90.

174. *Radio Listeners in Panic, Taking War Drama as Fact*, N.Y. TIMES, Oct. 31, 1938, at A1.

175. United States v. Brahm, 520 F. Supp. 2d 619, 627 (D.N.J. 2007).

176. 567 U.S. 709 (2012).

177. *Id.* at 722.

178. *Id.* at 730.

179. *Id.* at 709.

180. *Id.* at 714.

181. *Id.* at 720.

have been permitted for a few historic categories of speech, discussed in Section IV.B.2., *supra*, any general exclusion of protection for false statements had been absent from that group.[182] In Justice Kennedy's view, the Court had never endorsed the categorical rule that false statements receive no First Amendment protection.[183]

Without employing the term "strict scrutiny," the plurality moved from the categorical approach — under which the Stolen Valor Act did not fit into any existing exception to First Amendment protection — to the application of what is called "exacting scrutiny"[184] — to determine that a new category of unprotected speech should not be recognized. The government failed to establish a direct causal link between its compelling interest in protecting the integrity of the military honors system and the restriction on false speech.[185] Since counter-speech, through public refutation of a false claim, could vindicate the government's interests, the law was "not actually necessary."[186] Moreover, the availability of "less speech-restrictive" alternatives, such as a government database that listed Congressional Medal of Honor winners, enhanced the law's infirmity.[187] Invoking Justice Holmes, the plurality proclaimed that "[t]he remedy for speech that is false is speech that is true."[188] And Justice Kennedy espoused the merits of the marketplace of ideas:

> The First Amendment itself ensures the right to respond to speech we do not like, and for good reason. Freedom of speech and thought flows not from the beneficence of the state but from the inalienable rights of the person. And suppression of speech by the government can make exposure of falsity more difficult, not less so. Society has the right and civic duty to engage in open, dynamic, rational discourse. These ends are not well served when the government seeks to orchestrate public discussion through content-based mandates.[189]

The plurality maintained that no prior Court decision had confronted a measure like the Stolen Valor Act that targeted "falsity and nothing more."[190] It distinguished the Act from permissible laws that proscribe false speech, such as those prohibiting lying to government officials,

182. *Id.* at 718.
183. *Id.* at 719.
184. *Id.* at 715.
185. *Id.* at 725.
186. *Id.* at 726.
187. *Id.* at 729.
188. *Id.* at 727.
189. *Id.* at 728.
190. *Id.* at 709.

punishing perjury, or impersonating a government official, where the societal interest was beyond the prevention of the falsehood itself.[191] Unlike those laws, the Stolen Valor Act did not require an intent to cause harm or gain materially from the falsehood, giving it extraordinary reach: It applied "to a false statement made at any time, in any place, to any person."[192] If the government could criminalize this speech, the plurality reasoned, such a holding "would endorse government authority to compile a list of subjects about which false statements are punishable."[193]

Justice Breyer, joined by Justice Kagan, concurred in the judgment that the law violated the First Amendment. Foregoing categorical analysis, he instead applied intermediate scrutiny, concluding that the social benefits of the Act were disproportionate to its constitutional harm. When reviewing the constitutionality of a statute under the First Amendment, the Court, he wrote, "often found" it useful to apply what was sometimes called "intermediate scrutiny," "'proportionality' review" or "examination of 'fit.'"[194] While Justice Breyer's analysis — sometimes referred to as the "balancing method," for balancing free speech values against other societal interests on a case-by-case basis — has been rejected by a majority of the Court,[195] his weighing of the competing factors at issue[196] was essentially identical to the plurality opinion.[197]

Like the plurality, Justice Breyer concluded that few, if any, statutes simply prohibit the telling of a lie. He cited federal false reporting statutes as evidence of the proposition that "[s]tatutes prohibiting false claims of terrorist attacks, or other lies about the commission of crimes or catastrophes, require proof that substantial public harm be directly foreseeable, or, if not, involve false statements that are very likely to bring about that harm."[198] Limiting features justified other statutes and doctrines that punish the communication of false statements:

> [I]n virtually all these instances limitations of context, requirements of proof of injury, and the like, narrow

191. *Id.* at 710.

192. *Id.* at 722.

193. *Id.* at 723.

194. *Id.* at 730 (Breyer, J., concurring).

195. *See* SMOLLA, *supra* note 125, at § 2:11 (2017) (citing District of Columbia v. Heller, 554 U.S. 570 (2008)) ("In *Heller*, the Supreme Court's landmark Second Amendment decision, the Court majority, responding to the dissenting views of Justice Breyer, argued forcefully that 'interest balancing' fails to secure constitutional rights in the sense contemplated by our constitutional tradition.").

196. *Alvarez*, 567 U.S. at 730–32 (Breyer, J., concurring).

197. *Alvarez*, 567 U.S. at 711.

198. *Id.* at 735 (citing 47 C.F.R. § 73.1217 (2011) ("requiring showing of foreseeability and actual substantial harm"); 18 U.S.C. § 1038(a)(1) (2012) ("prohibiting knowing false statements claiming that terrorist attacks have taken, are taking, or will take, place").

> the statute to a subset of lies where specific harm is more likely to occur. The limitations help to make certain that the statute does not allow its threat of liability or criminal punishment to roam at large, discouraging or forbidding the telling of the lie in contexts where harm is unlikely or the need for the prohibition is small.[199]

Since the breadth of the Act created a significant risk of First Amendment harm, Justice Breyer held out the possibility that a more narrowly drawn statute "could significantly reduce the threat of First Amendment harm while permitting the statute to achieve its important protective objective."[200]

In dissent, three Justices voted to uphold the Act based on a narrower view of the protection that the Constitution affords lies. Justice Alito, joined by Justice Scalia and Justice Thomas, maintained that the Court's precedents "amply demonstrate that false statements of fact merit no First Amendment protection in their own right."[201] The dissent relied on the legislative determination that the false statements undermined the country's system of military honors and inflicted real harm on actual medal recipients and their families.[202] As "false factual statements that inflict real harm and serve no legitimate interest," then, the speech proscribed by the Act was unprotected — unless their prohibition would chill other expression that falls within the Amendment's scope.[203]

Still, all three opinions agreed that some lies warrant constitutional protection. The dissent accepted Justice Breyer's list of discrete categories of false statements that serve a valid purpose: false statements that "'prevent embarrassment, protect privacy, shield a person from prejudice, provide the sick with comfort, or preserve a child's innocence' . . . 'stop a panic or otherwise preserve calm in the face of danger' or further philosophical or scientific debate."[204]

In response to the Court's ruling, Congress passed, and President Obama signed, a new version of the Stolen Valor Act into law.[205] Following Justice Breyer's directive, the amended Stolen Valor Act of 2013[206] narrowed the reach of the statute by imposing a *mens rea* requirement. Specifically, the person telling the lie must now do so with

---

199. *Id.* at 736.
200. *Id.* at 739.
201. *Id.* at 748–49 (Alito, J., dissenting).
202. *Id.* at 742–43.
203. *Id.* at 739.
204. *Id.* at 733.
205. Lee Ferran, *Obama Signs Stolen Valor Act into Law*, ABC NEWS (June 3, 2013), http://abcnews.go.com/blogs/headlines/2013/06/obama-signs-stolen-valor-act-into-law/ [https://perma.cc/BXP5-886P].
206. 18 U.S.C. § 704 (2013).

the "intent to obtain money, property, or other tangible benefit" from the lie.[207]

Last year, in *United States v. Swisher*,[208] the Ninth Circuit, sitting *en banc*, used Justice Breyer's intermediate scrutiny test to strike down another provision of the Stolen Valor Act that criminalized false speech.[209] The court held that Section 704(a), which prohibited wearing an unauthorized military medal, was an unconstitutional content-based restriction on free speech.[210] The law failed the first prong of Justice Breyer's intermediate scrutiny test, which requires consideration of "the seriousness of the speech-related harm the provision will likely cause."[211] The Court determined that the law created a "significant risk of First Amendment harm" for the same reasons as the provision at issue in *Alvarez*: it required no act beyond the false communication itself, it had the same broad reach, and likewise did not require that a specified harm would result from the falsehood.[212] While the Court concluded that the government had the same compelling interest in enacting Section 704(a) as it did in enacting Section 704(b), satisfying Breyer's second prong, there existed both an insufficient causal link between that interest and the restriction and less restrictive ways of achieving said interest.[213] As explained in *Alvarez*, Congress could adopt narrowing strategies to limit the breadth of the prohibition, and could establish "information-disseminating devices," as equally effective means to meeting the government's goals.[214] Given that the provision failed the intermediate scrutiny test, it could not survive the plurality's exacting scrutiny test either.[215]

## V. First Amendment Challenges to False Reporting Statutes as Applied on Social Media

*Alvarez* serves as powerful support for a First Amendment challenge to New York's false reporting statute. Unlike the examples of narrowly tailored statutes described in the plurality and Justice Breyer's opinions, New York's statute does not require "proof that substantial public harm [is] directly foreseeable," nor does it require that the false statements be "very likely to bring about that harm."[216] Instead, the statute requires

207. *Id.*
208. 811 F.3d 299 (9th Cir. 2016).
209. *Id.* at 317–18.
210. *Id.*
211. *Id*. at 315 (quoting United States v. Alvarez, 567 U.S. 709, 730 (2012) (Breyer, J., concurring)).
212. *Swisher*, 811 F.3d at 315–16.
213. *Id*. at 317 (quoting *Alvarez*, 567 U.S. at 738 (Breyer, J., concurring)).
214. *Id.*
215. *Id*. at 317.
216. *Alvarez*, 567 U.S. at 735.

only a tenuous connection between the lie and the anticipated harm — i.e., that public alarm or inconvenience be "not unlikely" to result from the lie.[217]

Of course, there are many conceivable applications of the statute that do *not* raise constitutional problems, and many kinds of speech can be criminalized without difficulty. A person could be prosecuted, for example, for falsely reporting to a law enforcement officer an impending terrorist attack. But the constitutional application of the statute is suspect in other contexts, particularly in the context of false speech made on social media platforms. For that reason, this Article focuses on a First Amendment challenge to the statute as applied to false reports on social media.

To be sure, *Alvarez* involved a facial challenge, in which the plaintiff argued that no application of the statute would be constitutional.[218] But given *Alvarez*'s holding with regard to First Amendment protection for lies, generally, it is important precedent for analyzing the viability of a First Amendment challenge to a false reporting statute as applied to lies spread on social media, specifically.[219] The "as-applied" analysis below therefore relies, to some extent, on *Alvarez*.

### *A. Example: The Louisville "Purge" Hoax*

In 2014, a teenager in Louisville, Kentucky spread rumors on Twitter that his town was going to have a "purge," referring to the *Purge* films in which all crime is allowed for one night each year.[220] In response, the town cancelled a local football scrimmage and ordered additional police to patrol the streets.[221] There was no other response from the public or law enforcement.[222] Subsequently, the teen apologized and stated that he did not intend or expect anyone to panic as a result of his tweets.[223]

If the teen, whose name was not released, was charged under the New York false reporting statute, he might challenge the statute as applied to him under the First Amendment. The following discussion seeks

---

217. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).

218. Brief for Petitioner at 12, United States v. Alvarez, 567 U.S. 709 (2012) (No. 11-210).

219. In any event, the distinction between a facial challenge and as-applied challenge is perhaps less important than it may seem. As some have noted, "[r]eliance on ultimately superficial distinctions between facial and as applied challenges to statutes only confuses the underlying concerns of substantive constitutional doctrine and institutional competence that govern the resolution of each case." Michael C. Dorf, *Facial Challenges to State and Federal Statutes*, 46 STAN. L. REV. 235, 239 (1994).

220. Nicole Hensley, *Teenager Started 'Louisville Purge' Hoax*, DAILY NEWS (August 16, 2014), http://www.nydailynews.com/news/national/louisville-purge-hoax-started-teenager-police-article-1.1906062 [https://perma.cc/FAA4-5H5F].

221. *Id.*

222. *Id.*

223. *Id.*

to determine whether the statute would pass constitutional muster in such circumstances.

1. The New York False Reporting Statute Is a Content-Based Restriction on Speech

As a threshold matter, the teen would need to show that the New York statute restricts speech or expressive conduct in order for the First Amendment to apply.[224] On first blush, one might think that the act of "circulating" information via social media is non-expressive conduct, which is unprotected.[225] Legal precedent, however, suggests otherwise. As the North Carolina Supreme Court held when addressing Facebook posts that violated North Carolina's cyberbullying statute, "[s]uch communication does not lose protection merely because it involves the 'act' of posting information online, for much speech requires an 'act' of some variety . . . ."[226] Accordingly, the teen prankster's online post would likely be deemed to constitute "speech" entitled to First Amendment protection as a threshold matter.

Further, the New York statute would likely qualify as a content-based restriction. Just as in *Alvarez*, the New York statute's prohibition extends only to a certain type of speech (false speech) that spreads a certain message (the "occurrence or impending occurrence of a crime, catastrophe or emergency").[227] As such, the statute would likely be subject to strict scrutiny review — i.e., the law is presumed invalid, and the government must show that it is narrowly tailored to promote a compelling government interest.

The critical issues, then, are (1) what government interests are implicated for proscribing the dissemination of false reports, and (2) whether the false reporting statute is narrowly tailored to promote those interests in the context of speech made on social media platforms.

---

224. *See, e.g.*, Texas v. Johnson, 491 U.S. 397, 403 (1989).

225. *See, e.g.*, *R.A.V.*, 505 U.S. at 389 ("[W]ords can in some circumstances violate laws directed not against speech but against conduct (a law against treason, for example, is violated by telling the enemy the Nation's defense secrets) . . . ."); State v. Camp, 295 S.E.2d 766, 769 (1982) (opining that a statute barring use of a telephone to harass another person implicated conduct, not speech, and therefore did not violate the First Amendment).

226. State v. Bishop, 787 S.E.2d 814, 818 (N.C. 2016); *see also* Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 790, (2011) ("And whatever the challenges of applying the Constitution to ever-advancing technology, 'the basic principles of freedom of speech and the press, like the First Amendment's command, do not vary' when a new and different medium for communication appears." (quoting Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 503 (1952))).

227. United States v. Alvarez, 567 U.S. 709, 715 (2012) ("The Government contends the criminal prohibition is a proper means to further its purpose in creating and awarding the Medal. When content-based speech regulation is in question, however, exacting scrutiny is required.").

2. The Government Has a Compelling Interest to Restrict False Reports Because False Reports Cause Alarm and Waste Resources

The government must have a compelling interest to overcome strict scrutiny. Here, a court would likely find a compelling interest promoted by the New York statute because false reports could cause unnecessary alarm, unrest, and the diversion of emergency services. Indeed, the New York statute was enacted to augment offenses that proscribed giving false fire alarms and circulating false "bomb scare" reports[228], and would "guard against the inconvenience and alarm that may be occasioned by circulating a false alarm."[229]

The Supreme Court has previously upheld the "interest of the community in maintaining peace and order on the streets."[230] Although the Court subsequently held in several cases that community unrest was not a sufficient justification for restricting otherwise protected speech,[231] recent holdings in the anti-hoax context suggest that preventing unrest is more likely to be a compelling interest when the causal speech is objectively and knowingly false.

For example, in *United States. v. Brahm*,[232] the U.S. District Court for the District of New Jersey considered a First Amendment challenge to the federal anti-hoax statute, 18 U.S.C. § 1038.[233] The court found that "[t]he government interests protected by § 1038 are preservation of order and protection of emergency services personnel from wasteful and potentially risky responses to nonexistent threats."[234] In addition, the legislative history of the statute revealed concerns that hoaxes "aid terrorists, endanger public health, and instill fear into the public."[235] The

228. N.Y. PENAL LAW § 240.50(1) cmt. practice (McKinney 2016).

229. MODEL PENAL CODE § 250.3 cmt. 2 at 356–57 (AM. LAW INST. 1980).

230. Feiner v. New York, 340 U.S. 315, 320–21 (1951) (upholding a speaker's conviction for disorderly conduct and holding that communities may punish "when as here the speaker passes the bounds of argument or persuasion and undertakes incitement to riot").

*231. See, e.g.*, Henry v. City of Rock Hill, 376 U.S. 776, 778 (1964) (holding that communities cannot punish speakers simply because "their speech stirred people to anger, invited public dispute, or brought about a condition of unrest"); Cox v. Louisiana, 379 U.S. 536, 551 (1965) (invalidating a statute that made "breach of the peace" unlawful, which had been judicially defined as "to agitate, to arouse from a state of repose, to molest, to interrupt, to hinder, to disquiet"); Edwards v. South Carolina, 372 U.S. 229, 237 (1963) (holding that speech must be protected when, and is perhaps most useful when, "it induces a condition of unrest, creates dissatisfaction with conditions as they are, or even stirs people to anger").

232. 520 F. Supp. 2d 619 (D.N.J. 2007).

233. *Id.* at 622. Section 1038(a)(1) criminalizes "engag[ing] in any conduct with intent to convey false or misleading information under circumstances where such information may reasonably be believed and where such information indicates that an activity has taken, is taking, or will take place that would constitute a violation of [numerous predicate criminal acts involving, inter alia, nuclear, biological, or chemical weapons, transportation, buildings, and explosives]." 18 U.S.C. § 1038(a)(1) (2012).

234. *Brahm*, 520 F. Supp. 2d at 628.

235. *Id.*; H.R. Rep. No. 108-505, at 4 (2004).

district court upheld the statute after concluding that "[t]he state interest in these issues is very strong."[236]

Similarly, in *United States v. Keyser*, the Ninth Circuit analyzed Section 1038 and concluded that "[p]rompting law enforcement officials to devote unnecessary resources and causing citizens to fear they are victims of a potentially fatal terrorist attack is 'the sort of harm . . . Congress has a legitimate right to prevent by means of restricting speech.'"[237]

In light of this precedent, a court would likely find that the government has a compelling interest in preventing alarm and the waste of resources that may result from false speech about crimes, catastrophes, and emergencies. The analysis does not end there, however — the statute must be narrowly tailored to achieve the government interest to pass constitutional muster.[238]

3. The New York False Reporting Statute Is Not Narrowly Tailored to Promote the Government's Interest

New York's false reporting statute would likely be held impermissibly broad as applied to the Louisville prankster because it criminalizes and chills speech that results in only minor public alarm and inconvenience. Specifically, the New York statute requires no act beyond the communication of a falsehood, no intent to cause harm, and no actual harm.[239] All that is required is that the speaker intend to communicate the falsehood, and that the falsehood be *not unlikely* to cause harm (i.e., "public alarm" or "inconvenience"). In addition, there are less restrictive alternatives available for deterring and detecting false speech on social media, making the law not "actually necessary" as is required.

*a. The Statute Proscribes No Act Beyond the Communication Itself*

First, the statute proscribes the mere act of communication. In *Alvarez*, one reason that the Supreme Court struck down the Stolen Valor Act

236. *Brahm*, 520 F. Supp. 2d at 628.

237. United States v. Keyser, 704 F.3d 631, 640 (9th Cir. 2012) (citing United States v. Alvarez, 617 F.3d 1198, 1215 (9th Cir. 2010)); *see also* United States v. Alvarez, 567 U.S. 709, 732 (2012) (Breyer, J., concurring) ("The dangers of suppressing valuable ideas are lower where, as here, the regulations concern false statements about easily verifiable facts . . . Such false factual statements are less likely than true factual statements to make a valuable contribution to the marketplace of ideas.").

238. *Alvarez*, 567 U.S. at 725 (2012) ("But to recite the Government's compelling interests is not to end the matter. The First Amendment requires that the Government's chosen restriction on the speech at issue be actually necessary to achieve the interest." (citing Brown v. Entm't Merchs. Ass'n, 564 U.S. 786, 799, (2011)).

239. *See Alvarez*, 567 U.S. at 734 (explaining that constitutional restrictions on false speech require either "proof of specific harm to identifiable victims," or a particularly great likelihood of harm).

was because it criminalized false speech "made at any time, in any place, to any person . . . whether shouted from the rooftops or made in a barely audible whisper."[240] Applying that reasoning to the New York statute reveals a serious First Amendment concern.

The only conduct required to violate the New York statute is the knowing "initiat[ion] or circulat[ion]" of a false report.[241] The government need not prove that the speaker made the report to any particular person or agency. Furthermore, the message could be sent to just one person as long as the message is "not unlikely" to cause alarm or inconvenience. In this regard, the New York statute is similar to the law struck down in *People v. Marquan*.[242] There, the New York Court of Appeals analyzed the constitutionality of a cyberbullying statute that criminalized "any act of communicating or causing a communication to be sent by mechanical or electronic means . . . with the intent to harass, annoy, threaten, abuse, taunt, intimidate, torment, humiliate, or otherwise inflict significant emotional harm on another person."[243] The court struck down the law despite the government's compelling interest (and laudable goal) in "protecting children from harmful publications or materials."[244] The court concluded that the law had "alarming breadth" because it "would criminalize a broad spectrum of speech outside the popular understanding of cyberbullying."[245]

New York's false reporting statute similarly criminalizes any falsehood regarding the alleged occurrence or impending occurrence of a crime, catastrophe, or emergency with the potential to cause harm. In the case of the "purge" hoax, a court could find that public alarm and inconvenience was "not unlikely" to result from the tweet, because the *Purge* films are heavily advertised and known to at least some segments of the movie-going public. Those people, recognizing that "purge" refers to unchecked crime and mayhem, could be alarmed and act accordingly. Thus, under the New York statute's broad and nebulous "not unlikely" standard, the prankster could be convicted simply for his speech even if the post did not actually result in any harm. *Alvarez* and *Marquan* suggest that such proscriptions of false speech are unconstitutionally broad.[246]

240. *Id.* at 722–23.
241. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).
242. *See* People v. Marquan, 19 N.E.3d 480, 484 (N.Y. 2014).
243. *Id.*
244. *Id*. at 485.
245. *Marquan*, 19 N.E.3d at 486.
246. *See generally* United States v. Alvarez, 567 U.S. 709 (2012); *Marquan*, 19 N.E.3d 480.

*b. The Statute Does Not Require Intent to Cause Harm*

Second, the statute lacks an intent requirement. Concurring in *Alvarez*, Justice Breyer interpreted the Stolen Valor Act to require knowledge of the falsehood *and intent that the false information be taken as true*, but rejected the law nonetheless because, "although this interpretation diminishes the extent to which the statute endangers First Amendment values, it does not eliminate the threat."[247] Justice Breyer noted that this threat to speech was especially dangerous because "false factual statements can serve useful human objectives . . . they may shield a person from prejudice . . . they may stop a panic or otherwise preserve calm in the face of danger."[248]

The only *mens rea* requirement imposed by the New York statute, by contrast, is "knowing the information reported, conveyed or circulated [is] false or baseless."[249] Therefore, the speaker need not intend to cause alarm or inconvenience to be guilty of a crime.[250] In the "purge" hoax example, the teen knew that his government was not going to allow all crime for one night. His knowledge alone would subject him to liability under the New York statute, even though he had no intent for his statement to be taken as true and no intent to cause any particular result, let alone harm to any person. Such broad criminal liability flies in the face of *Alvarez*.[251]

Because it is lacking an intent requirement, the New York statute is also different from the Model Penal Code's false reporting provision.[252] Under the Code, a violator must "know that his conduct is 'likely to cause evacuation of a building, place of assembly, or facility of public transport, or to cause public inconvenience or alarm.'"[253] Thus, excluded from liability is "the practical joker or other person who circulates a false alarm in circumstances where he is unaware of the potential for serious consequences."[254] The teen prankster falls squarely in this category of persons excluded from liability in the Model Penal Code but subject to liability under the New York statute.

---

247. *Alvarez*, 567 U.S. at 732 (Breyer, J., concurring); *see also* State v. Bishop, 787 S.E.2d 814, 819–22 (N.C. 2016) (striking down a cyberbullying law even though it required "intent to intimidate or torment" because the terms swept in essentially harmless speech).

248. *Alvarez*, 567 U.S. at 733 (Breyer, J., concurring).

249. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).

250. In contrast to the federal anti-hoax statute, which requires "intent to convey false or misleading information," 18 U.S.C. § 1038 (2012).

251. *See Alvarez*, 617 F.3d at 1200 (9th Cir. 2010) ("All previous circumstances in which lies have been found proscribable involve not just knowing falsity . . . . Indeed, if the Act is constitutional . . . , then there would be no constitutional bar to criminalizing lying about one's height, weight, age, or financial status on Match.com or Facebook . . . .").

252. MODEL PENAL CODE § 250.3 (AM. LAW INST. 1980).

253. *Id.*

254. MODEL PENAL CODE § 250.3 cmt. 2 (AM. LAW INST. 1980).

In a case like the teen prankster's, false speech may be humor or satire, which have been recognized as legitimate and important forms of speech.[255] The Stolen Valor Act was injurious to free speech because it proscribed useful false statements along with more harmful ones. By not requiring any intent beyond knowledge, the New York statute similarly imposes a risk of liability on speakers whose speech is knowingly false but intended as self-expression, political commentary, or entertainment. Despite the legitimate purpose of the statute, this is a risk that likely cannot withstand constitutional scrutiny.[256]

*c. The Statute Does Not Require Actual Harm*

Third, the statute does not require harm to result from the false report to impose liability. A speaker can be guilty under the New York statute without causing any harm, or without harm even being a likely result. The statute requires only that the falsehood be made "under circumstances in which it is not unlikely that public harm or inconvenience will result[.]"[257] This makes the law significantly broader than constitutional restrictions on false speech, which are typically constrained by "requiring proof of specific harm to identifiable individuals," or "limiting the prohibited lies to those that are particularly likely to produce harm."[258]

The incitement doctrine provides useful guidance on what speech restrictions are and are not permissible. In 1919, the Supreme Court held that speech may be restricted as long as there is a "clear and present danger" that it will produce harm.[259] By 1969, the Court had largely abandoned the clear and present danger test and had arrived at a more

255. *See* Robert D. Sack, *Sack on Defamation*, 116 (4th ed. 2010) ("Humor is an important medium of legitimate expression and central to the well-being of individuals, society, and their government. Despite its typical literal 'falsity,' any effort to control it runs severe risks to free expression as dangerous as those addressed to more 'serious' forms of communication."); *see also* Michael A. Einhorn, *Miss Scarlett's License Done Gone!: Parody, Satire, and Markets*, 20 CARDOZO ARTS & ENT. L.J. 589, 603 (2002) (contending that satire is useful because it uses recognized symbols to "ridicule or criticize political institutions, cultural values, or media presentations").

256. *See* United States v. Stevens, 559 U.S. 460, 460–61 (2015) ("[T]he First Amendment's guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits."); *Alvarez*, 567 U.S. at 723 ("Were the Court to hold that the interest in truthful disclosure alone is sufficient to sustain a ban on speech, absent any evidence that the speech was used to gain a material advantage, it would give government broad censorial power unprecedented in this Court's cases or in our constitutional tradition.").

257. N.Y. PENAL LAW § 240.50(1) (2013).

258. *See Alvarez*, 567 U.S. at 734 (Breyer, J., concurring) (explaining how specific constitutionally valid statutes have limits to their restrictions on false speech); *see also* United States v. Williams, 690 F.3d 1056, 1063 (8th Cir. 2012) (finding anti-threat statutes permissible in light of *Alvarez* because the statutes "criminalize only those lies that are particularly likely to produce harm").

259. Schenck v. United States, 249 U.S. 47, 52 (1919).

exacting test: speech can now qualify as incitement, and thus be restricted without violating the First Amendment, only if it is (1) directed to inciting or producing imminent lawless action and (2) likely to incite or produce such action.[260] The dual requirements of intent and a likelihood of harm in the incitement context are in stark contrast to New York's restriction on false speech, which requires *neither* intent nor a likelihood of harm.

A recent case in the cyberbullying context is also instructive. In *State v. Bishop*, the North Carolina Supreme Court heard a challenge to a law under which it was "unlawful for any person to use a computer or computer network to . . . [p]ost or encourage others to post on the Internet private, personal, or sexual information pertaining to a minor . . . [w]ith the intent to intimidate or torment a minor."[261] The court struck down the law because, "[e]ven under the State's interpretation of [the statute], the statute prohibits a wide range of online speech — whether on subjects of merely puerile interest or on matters of public importance — and all with no requirement that anyone suffer any actual injury."[262]

Similar to the law considered in *Bishop*, the New York statute potentially criminalizes online speech of puerile interest — for example, a hoax based on popular (but unrealistic) horror films — and speech on matters of public concern (e.g., the state of affairs during an emergency), with no requirement that anyone suffer actual injury. In fact, the New York law is even broader than the cyberbullying statute at issue in *Bishop* because it lacks an intent requirement. Therefore, if upheld, the New York statute could criminalize the Louisville teen's speech even though the actual harm caused by the speech was only the cancellation of a local football game. This is likely too restrictive of the First Amendment's protection.

### *d. Less Restrictive Alternatives Exist to Deter and Detect False Speech on Social Media*

Fourth, there are less restrictive alternatives available to the government for combatting false speech likely to cause public alarm on social media. One of the most problematic aspects of the Stolen Valor Act struck down in *Alvarez* was that "[t]he Government ha[d] not shown, and [could not] show, why counterspeech would not suffice to achieve its interest … the dynamics of free speech, of counterspeech, of refuta-

260. Brandenburg v. Ohio, 395 U.S. 444, 447 (1969); *see also* United States v. Williams, 553 U.S. 285, 322 (2008) 2 (Souter, J., dissenting) ("*Brandenburg* unmistakably insists that any limitation on speech be grounded in a realistic, factual assessment of harm.").

261. *Bishop*, 787 S.E.2d at 815 (citing N.C. GEN. STAT. ANN. § 14-458.1(a)(1)(d)).

262. *Bishop*, 787 S.E.2d at 821.

tion, can overcome the lie."[263] Simply put, "[t]he remedy for speech that is false is speech that is true."[264]

Similarly, counterspeech would likely be sufficient to combat false speech on social media. As discussed above, social media platforms are information-disseminating fora. By the very nature of social media, falsehoods can quickly and effectively be countered by truth, making the criminalizing of false speech on social media not "actually necessary" to prevent alarm and inconvenience. As described above, in the wake of the Boston Marathon bombing, there was a good deal of false information spreading on various social media platforms.[265] But using those very same platforms, the Boston Police Department quickly refuted and corrected the misinformation. The BPD tweeted an accurate casualty number in response to inflated reports, refuted rumors that a fire at the John F. Kennedy Presidential Library was related to the bombing, and corrected another rumor that a Saudi man had been arrested.[266]

In addition to individuals and organizations posting corrections, increasing numbers of websites and technologies exist to help prevent the spread of misinformation online. One popular website, for example, is snopes.com, where people can fact-check rumors.[267] Additionally, researchers in Qatar and India released a program called TweetCred that rates the credibility of Twitter posts in real time.[268] A business consultant and a developer in Germany launched Hoaxmap, an online platform aimed at debunking false rumors.[269] And in the United States, patents have been issued for methods and systems for detecting lies on social media.[270]

The "purge" hoax could easily be, and in fact *was*, refuted by the counterspeech of other social media users. Because true speech is an available and effective remedy, the New York statute is not the least restrictive alternative for limiting such speech on social media. The law would therefore likely fail strict scrutiny as applied to the teen, because it is not "actually necessary."

Thus, for these reasons, broad false reporting statutes like New York's may be susceptible to a First Amendment challenge as applied to

---

263. *Alvarez*, 567 U.S. at 726.

264. *Id.* at 727.

265. *See* Reinwald, *supra* note 10.

266. *See* DAVIS III ET AL., *supra* note 30.

267. *See, e.g.*, David Emery, *Instant Replay: A video allegedly showing anti-Trump protesters beating a man to death in Philadelphia was actually filmed in 2014 and is unrelated to the protests*, SNOPES (Nov. 12, 2016), http://www.snopes.com/protesters-beat-homeless-veteran/ [http://perma.cc/9LXB-3DAN].

268. Dewey, *supra* note 76.

269. Federico Guerrini, *Using Crowdsourcing to Debunk Social Media Hoaxes about Refugee Crimes in Germany*, FORBES (Feb. 22, 2016), http://www.forbes.com/sites/federicoguerrini/2016/02/22/brilliant-crowdsourced-hoaxmap-helps-debunk-false-reports-about-refugee-crimes/#3f19a3e57dba [https://perma.cc/Y3SA-SEC8].

270. *See* U.S. Patent Nos. 9,361,382, and 9,047,253.

false speech on social media, at least in the circumstances described here.

### *B. False Reporting Statutes as Applied to Social Media Pose a Significant Threat of First Amendment Harm*

Beyond the statutory analysis, there are policy concerns that support such a conclusion. Putting aside the examples discussed above in which the speakers *knew* that their speech was false, statutes like New York's are likely to cast a chill on speech that the speaker *thinks* is true, regardless of whether it is.[271] When the only *mens rea* requirement is knowledge, and when "inconvenience" need only be "not unlikely" to result,[272] people may refrain from engaging in protected speech for fear of legal penalties.[273] The Court in *Alvarez* expressed a similar concern, noting that "the pervasiveness of false statements, made for better or for worse motives, made thoughtlessly or deliberately, made with or without accompanying harm, provides a weapon to a government broadly empowered to prosecute falsity without more."[274] Further, because of the ubiquitous nature of falsehoods on social media and the consequent inability of the government to prosecute each possible violation of a false reporting statute, "those who are unpopular may fear that the government will use that weapon selectively."[275] In other words, people may refrain from making statements that they *believe* to be true, for fear that the statement will turn out to be false and they will be unable to refute the government's claim that they *knew* it was false.

The believed-to-be-true speech chilled by false reporting statutes is not only constitutionally protected, it also has social utility. As discussed above, in the hours after the Boston Marathon bombings, the BPD used Twitter to request public assistance and to keep the public and the media informed about the casualty toll and the status of the investigation.[276] When news operations like *The Huffington Post* and *BuzzFeed* lost use

---

271. *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974) (explaining that the threat of criminal prosecution for making a false statement can inhibit the speaker from making true statements, thereby chilling a kind of speech that lies at the First Amendment's heart).

272. N.Y. PENAL LAW § 240.50(1) (McKinney 2016).

273. *See Alvarez*, 567 U.S. at 736 (Breyer, J., concurring) ("[A] speaker might still be worried about being prosecuted for a careless false statement, even if he does not have the intent required to render him liable. And so the prohibition may be applied where it should not be applied, for example, to bar stool braggadocio."); *see also Bishop*, 787 S.E.2d at 821 ("Were we to adopt the State's position, it could be unlawful to post on the Internet any information "relating to a particular [minor]." Such an interpretation would essentially criminalize posting any information about any specific minor if done with the requisite intent . . . behavior that a robust contemporary society must tolerate because of the First Amendment, even if we do not approve of the behavior.").

274. *Alvarez*, 567 U.S. at 734 (Breyer, J., concurring).

275. *Id.*

276. *See* DAVIS III ET AL., *supra* note 30.

of their servers during Hurricane Sandy, they, too, turned to Twitter and other social media to deliver reports.[277] And in the hours before Hurricane Gustav arrived in New Orleans, at least one person was persuaded to evacuate, not by news reports, but by the number of friends on Twitter reporting that they were evacuating.[278] It is unlikely that any of these sources *knew* with certainty that their reports were accurate, but they spoke with the intention of keeping others informed and safe. As Justice Breyer pointed out, ultimately false speech is the price we pay for speech that has the potential to preserve calm and shield people from prejudice during periods of unrest.[279] This is especially true in the ubiquitous context of social media. Yet, if people fear that they will later be punished for inconvenient responses to their well-intentioned speech, they may not speak at all.

Ultimately, broad false reporting statutes like the one in New York may counterproductively restrict one of the most powerful tools of informing and reassuring the public.[280]

## VI. CONCLUSION

In the social media age, false reports about emergencies have the potential to cause a great deal of public alarm and unrest. Governments therefore have an interest in deterring these false reports; however, some laws that impose liability for false speech are too broad. New York's false reporting statute is a prime example. The statute may be susceptible to a First Amendment challenge, at least as applied to falsehoods made on social media. In light of the Supreme Court's decision in *Alvarez*, the statute is problematic in its proscription of mere lies without requiring intent to cause harm. The existence of less restrictive alternatives to combat false reports suggests that such statutes are not "actually necessary" as required by the First Amendment.

Social media provides an accessible forum that allows anyone to publish speech that will reach millions of people around the world with just a click of a mouse. The immediacy and pseudonymous nature of social media has bred a more cavalier attitude toward truth over the past several years, resulting in the widespread dissemination of false information about newsworthy events, including emergencies and natural catastrophes. As a result, social media is now rife with false speech, and it is only a matter of time before those who use social media to spread

277. Guskin & Hitlin, *supra* note 34.

278. *See generally* Mills et al., *supra* note 35 (providing numerous examples of social media's positive influence during emergencies, and concluding that social media has better quality information than mainstream media within the first 24 hours of an emergency).

279. *Alvarez*, 567 U.S. at 733 (Breyer, J., concurring).

280. *See id.* at 728 ("And suppression of speech by the government can make exposure of falsity more difficult, not less so.").

misinformation are prosecuted in the United States. If that were to happen, a First Amendment challenge will be inevitable.