THE YALE LAW JOURNAL FORUM
OCTOBER 9, 2017

Real Talk About Fake News: Towards a Better Theory for Platform Governance

Nabiha Syed

Following the 2016 U.S. presidential election, "fake news" has dominated popular dialogue and is increasingly perceived as a unique threat to an informed democracy. Despite the common use of the term, it eludes common definition.[1] One frequent refrain is that fake news—construed as propaganda, misinformation, or conspiracy theories—has always existed,[2] and therefore requires no new consideration. In some ways this is true: tabloids have long hawked alien baby photos and Elvis sightings. When we agonize over the fake news phenomenon, though, we are not talking about these kinds of fabricated stories.

Instead, what we are *really* focusing on is why we have been suddenly inundated by false information—purposefully deployed—that spreads so quickly and persuades so effectively. This is a different conception of fake news, and it presents a question about how information operates at scale in the internet era. And yet, too often we analyze the problem of fake news by focusing on indi-

1. Claire Wardle, *Fake News. It's Complicated*, FIRST DRAFT (Feb. 16, 2017), http://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79 [http://perma.cc/EJ9Y-EP6V].

2. *See*, *e.g.*, Luciano Floridi, *Fake News and a 400-Year-Old Problem: We Need to Resolve the 'Post-Truth' Crisis*, GUARDIAN (Nov. 29, 2016, 12:42 AM), http://www.theguardian.com/technology/2016/nov/29/fake-news-echo-chamber-ethics-infosphere-internet-digital [http://perma.cc/X74P-7GUZ]; Arianna Huffington & Ari Emanuel, *Fake News: A New Name for an Old Problem*, HUFFINGTON POST (Dec. 21, 2016, 2:03 PM), http://www.huffingtonpost.com/entry/fake-news-a-new-name-for-an-old-problem_us_585acd94e4b0eb586484eab2 [http://perma.cc/EH3Y-DBFA]; Christopher Woolf, *Back in the 1890s, Fake News Helped Start a War*, PRI (Dec. 8, 2016), http://www.pri.org/stories/2016-12-08/long-and-tawdry-history-yellow-journalism-america [http://perma.cc/QZ63-7B3H].

vidual instances,[3] not systemic features of the information economy. We compound the problem by telling ourselves idealistic, unrealistic stories about how truth emerges from online discussion. This theoretical incoherence tracks traditional First Amendment theories, but leaves both users and social media platforms ill-equipped to deal with rapidly evolving problems like fake news.

This rupture gives us an excellent opportunity to reexamine whether existing First Amendment theories adequately explain the digital public sphere. This Essay proceeds in three Parts: Part I briefly outlines how social media platforms have relied piecemeal on three discrete theories justifying the First Amendment—the marketplace of ideas, autonomy and liberty, and collectivist views—and why that reliance leaves platforms ill-equipped to tackle a problem like fake news. Part II then takes a descriptive look at several features that better describe the system of speech online, and how the manipulation of each feature affects the problem of misinformation. Finally, Part III concludes with the recommendation that we must build a realistic theory—based on observations as well as interdisciplinary insights—to explain the governance of private companies who maintain our public sphere in the internet era.

I. MOVING BEYOND THE MARKETPLACE

As a doctrinal matter, the First Amendment restricts government censorship, but as a social matter, it signifies even more.[4] As colloquially invoked, the "First Amendment" channels a set of commonly held values that are foundational to our social practices around free speech. When, for example, individuals incorrectly identify criticism as "violating First Amendment rights," they actually seek to articulate a set of values crucial to the public sphere, including the ability to express and share views in society.[5] The First Amendment shapes how we imagine desirable and undesirable speech. So conceived, it becomes clear that our courts are not the only place where the First Amendment comes to life.

---

3. *See*, *e.g.*, James Alefantis, *What Happened When 'Pizzagate' Came to My Restaurant*, WASH. POST (Apr. 20, 2017), http://www.washingtonpost.com/opinions/pizzagate-taught-us-the-value-of-community/2017/04/19/92cae67c-23b0-11e7-bb9d-8cd6118e1409_story.html [http://perma.cc/65R7-5R8F] (discussing the 'Pizzagate' hoax).

4. *See generally* Jack M. Balkin, *The First Amendment is an Information Policy*, 41 HOFSTRA L. REV. 1 (2012) (analyzing the connection between the First Amendment as a governmental constraint and as posing requirements on an infrastructure of free expression).

5. *Cf.* Jack M. Balkin, Commentary, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 26-28 (2004) (arguing that "[f]reedom of speech is becoming a generalized right against economic regulation of the information industries").

One implication of this understanding is that First Amendment theory casts a long shadow, which even private communications platforms[6]—like Facebook, Twitter, and YouTube—cannot escape. Internet law scholar Kate Klonick deftly illustrates how these three private platforms should be understood as self-regulating private entities, governing speech through content moderation policies:

> A common theme exists in all three of these platforms' histories: American lawyers trained and acculturated in First Amendment law oversaw the development of company content moderation policy. Though they might not have 'directly imported First Amendment doctrine,' the normative background in free speech had a direct impact on how they structured their policies.[7]

But First Amendment thinking comes in several flavors. Which of these visions of the First Amendment have platforms embraced?

A. Existing First Amendment Theories

Three First Amendment theories predominate: the marketplace of ideas, autonomy, and collectivist theories. However, as this Section demonstrates, none of these fully captures online speech.

One option is the talismanic "marketplace of ideas." Recognized as the "theory of our Constitution," the marketplace metaphor imagines that robust engagement with a panoply of ideas yields the discovery of truth—

---

6. Since they do not implicate government action, private communications platforms like Facebook, Twitter, Reddit, and YouTube are not as clearly bound by First Amendment doctrine as their predecessors might have been. To the contrary, these platforms enjoy broad immunity from liability based on the user-generated messages, photographs, and videos that populate their pages: Section 230 of the Communications Decency Act has long given them wide berth to construct their platforms as they please. 47 U.S.C. § 230 (2012). The purpose of this grant of immunity was both to encourage platforms to be "Good Samaritans" and take an active role in removing offensive content, but also to avoid free speech problems of collateral censorship. *See* Zeran v. Am. Online, Inc. 129 F.3d 327, 330-31 (4th Cir. 1997) (discussing the purposes of intermediary immunity in Section 230 as not only to incentivize platforms to remove indecent content, but also to protect the free speech of platform users).

7. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. (forthcoming 2017) (manuscript at 26), http://ssrn.com/abstract=2937985 [http://perma.cc/3L5D-96XQ].

eventually.[8] "More speech" should be the corrective to bad speech like falsehoods.[9] This vision predictably tilts away from regulation, on the logic that intervention would harm the marketplace's natural and dynamic progression.[10] That progression involves ideas 'competing' in the marketplace, a conception with two fundamental shortcomings, each relevant in an era of too much available information: What happens when individuals do not interact with contrary ideas because they are easy to avoid? And what happens when ideas are not heard at all because there are too many?

The marketplace also does not neatly address questions of power, newly relevant in the internet era. The marketplace metaphor sprang forth at a time when the power to reach the general population through "more speech" was confined to a fairly homogenous, powerful few. Individuals may have had their own fiefdoms of information—a pulpit, a pamphlet—but communicating to the masses was unattainable to most. Accordingly, the marketplace never needed to address power differentials when only the powerful had the technology to speak at scale. The internet, and particularly social media platforms, have radically improved the capabilities of many to speak, but the marketplace theory has not adjusted. For example, how might the marketplace theory address powerful speakers who drown out other voices, like Saudi Arabian "cyber troops" who flood Twitter posts critical of the regime with unrelated content

---

8. Justice Oliver Wendell Holmes first infused this view of free speech into Supreme Court jurisprudence:

> [W]hen men have realized that time has upset many fighting faiths, they may come to believe even more than they believe the foundations of their own conduct that the ultimate good desired is better reached by free trade in ideas—that the best test of truth is the power of the thought to get itself accepted in the competition of the market . . . .

Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting); *see also* United States v. Alvarez, 567 U.S. 709, 728 (2012) (plurality opinion) (describing Justice Holmes' quotation from *Abrams v. United States* as "the theory of our Constitution," and concluding that our "[s]ociety has the right and civic duty to engage in open, dynamic, rational discourse"); Citizens Against Rent Control v. City of Berkeley, 454 U.S. 290, 295 (1981) ("The Court has long viewed the First Amendment as protecting a marketplace for the clash of different views and conflicting ideas. That concept has been stated and restated almost since the Constitution was drafted."); Red Lion Broad. Co. v. FCC, 395 U.S. 367, 390 (1969) ("It is the purpose of the First Amendment to preserve an uninhibited marketplace of ideas in which truth will ultimately prevail . . . .").

9. Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring) ("[T]he remedy to be applied is more speech, not enforced silence. Only an emergency can justify repression.").

10. *See* Davis v. FEC, 554 U.S. 724, 755-56 (2008) (Stevens, J., concurring and dissenting in part) ("It is the purpose of the First Amendment to preserve an uninhibited marketplace of ideas in which truth will ultimately prevail." (quoting *Red Lion*, 395 U.S. at 390).

and hashtags to obscure the offending post?[11] As adopted by the platforms, the marketplace theory offers no answer. Put differently, the marketplace-as-platform theory only erects a building; there are no rules for how to behave once inside. This theory yields little helpful insight for a problem like fake news or other undesirable speech.

A second, related vision explains First Amendment values through the lens of individual liberty.[12] What counts here is only the "fundamental rule" that "a speaker has the autonomy to choose the content of his own message" because speech is a necessary exercise of personal agency.[13] All that matters is that one can express herself. Naturally, this theory also creates a strong presumption against centralized interference with speech.[14] While certainly enticing—and conveniently neutral for social media platforms interested in building a large user base—this theory is piecemeal. Focusing *only* on the self-expressive rights of the singular speaker offers no consideration of whether that speech is actually heard. It posits no process through which truth emerges from cacophony. In fact, it is not clear that fake news, as an articulation of one's self-expression, would even register as a problem under this theory.

---

11. Brian Whitaker, *How Twitter Robots Spam Critics of Saudi Arabia*, AL-BAB (July 28, 2016), http://al-bab.com/blog/2016/07/how-twitter-robots-spam-critics-saudi-arabia [http://perma.cc/5NW7-TRRD]; *see also* Samantha Bradshaw & Philip N. Howard, *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation* (Computational Propaganda Research Project, Working Paper No. 2017.12, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf [http://perma.cc/KV48-B2Z4] (discussing how government, military or political party teams, "cyber troops," manipulate public opinion over social media).

12. *See, e.g.*, C. Edwin Baker, *Autonomy and Free Speech*, 27 CONST. COMMENT 251, 259 (2011) (asserting that the "most appealing" theory of the First Amendment regards "the constitutional status of free speech as required respect for a person's autonomy in her speech choices"). *But see* Owen M. Fiss, *Why the State?*, 100 HARV. L. REV. 781, 785 (1987) (arguing that the First Amendment protects autonomy as a means of encouraging public debate, rather than as an end in itself).

13. Hurley v. Irish-American Gay, Lesbian, and Bisexual Group, 515 U.S. 557, 573 (1995). In contrast, the First Amendment provides minimal protection for the autonomy interests of a speaker who is engaged in commercial speech. Such speakers do not engage in a form of self-expression when they provide the public with information about their products and services. *See* C. Edwin Baker, *Scope of the First Amendment Freedom of Speech*, 25 UCLA L. REV. 964, 996 (1978); Martin H. Redish, *The Value of Free Speech*, 130 U. PA. L. REV. 591, 593 (1982); David A. J. Richards, *Free Speech and Obscenity Law: Toward a Moral Theory of the First Amendment*, 123 U. PA. L. REV. 45, 62 (1974).

14. *See* Fiss, *supra* note 12, at 785.

Third, and far less fashionable, is the idea that the First Amendment exists to promote a system of political engagement.[15] This "collectivist," or republican, vision of the First Amendment considers more fully the rights of citizens to *receive* information as well as the rights of speakers to express themselves. Practically and historically, this has meant a focus on improving democratic deliberation: for example, requiring that broadcasters present controversial issues of public importance in a balanced way, or targeting media oligopolies that could bias the populace. This theory devotes proactive attention to the full system of speech.[16]

The republican theory, which accounts for both listeners and speakers, offers an appealingly complete approach. The decreased costs of creating, sharing, and broadcasting information online means that everyone can be both a listener and a speaker, often simultaneously, and so a system-oriented focus seems appropriate. But the collectivist vision, like the marketplace and autonomy approaches, is still cramped in its own way. The internet—replete with scatological jokes and Prince cover songs—involves much more than political deliberation.[17] And so any theory of speech that focuses only on political outcomes will fail because it cannot fully capture *what actually happens* on the internet.

B. Which First Amendment Vision Best Explains Online Speech?

Online speech platforms—bound by neither doctrine nor by any underlying theories—have, in practice, fused all three of these visions together.

At their inception, many platforms echoed a libertarian, content-neutral ethos in keeping with the marketplace and autonomy theories. For example, Twitter long ago declared itself to be the "free speech wing of the free speech party," straying away from policing user content except in limited and extreme

---

15. *See, e.g.*, Robert Post, *Reconciling Theory and Doctrine in First Amendment Jurisprudence*, 88 CALIF. L. REV. 2353, 2362 (2000) (stating that "[t]he democratic theory of the First Amendment . . . protects speech insofar as it is required by the practice of self-government").

16. In *Red Lion Broadcasting Co. v. Federal Communications Commission*, 395 U.S. 367 (1969), a rare court-endorsed example of this thinking, the Supreme Court upheld rules requiring that the public receive fair coverage of public importance from television broadcasters. *See* Owen M. Fiss, *Free Speech and Social Structure*, 71 IOWA L. REV. 1405, 1409-10 (1986); Fiss, *supra* note 12, at 782; *see also* Robert Post, *The Constitutional Status of Commercial Speech*, 48 UCLA L. REV. 1, 7 (2000) (providing background on the public discourse theory).

17. Balkin, *supra* note 5, at 34 ("The populist nature of freedom of speech, its creativity, its interactivity, its importance for community and self-formation, all suggest that a theory of freedom of speech centered around government and democratic deliberation about public issues is far too limited.").

circumstances.[18] Reddit similarly positions itself as a "free speech site with very few exceptions," allowing communities to determine their own approaches to offensive content.[19] Mark Zuckerberg's argument that "Facebook is in the business of letting people share stuff they are interested in" presents an autonomy argument if ever there were one.[20] In the wake of the 2015 Charlie Hebdo attacks in Paris, Zuckerberg specifically vowed not to bow to demands to censor Facebook, [21] and then did so again in 2016, when he explained to American conservative leaders that the Facebook platform was "a platform for all ideas."[22] Taken at face value, platforms offer little recourse in response to undesirable speech like hate speech or fake news.

Platforms have also, however, long invoked the language of engagement, albeit not political engagement. Platforms have long governed speech through

---

18. Sarah Jeong, *The History of Twitter's Rules*, MOTHERBOARD (Jan. 14, 2016, 10:00 AM), http://motherboard.vice.com/read/the-history-of-twitters-rules [http://perma.cc/J843-X3VA]; *accord* Josh Halliday, T*witter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party*,' GUARDIAN (Mar. 22, 2012, 11:57 AM), http://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech [http://perma.cc/H3JF-94MT]. *But see* Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, VERGE (Apr. 13, 2016), http://www.theverge.com/2016/4/13/11387934/internetmoderator-history-youtube-facebook-reddit-censorship-free-speech [http://perma.cc/22X8-3MRY] (discussing Twitter's efforts to moderate its content more aggressively).

19. Jeff Stone, *Erik Martin Leaves Reddit Amid Debate Over Free Speech*, INT'L BUS. TIMES (Oct. 13, 2014, 3:03 PM), http://www.ibtimes.com/erik-martin-leaves-reddit-amid-debate-over-free-speech-1703954 [http://perma.cc/Q682-WYZA]; *see also* Bryan Menegus, *Reddit Is Tearing Itself Apart*, GIZMODO (Nov. 29, 2016), http://gizmodo.com/reddit-is-tearing-itself-apart-1789406294 [http://perma.cc/5AD6-WANZ] (describing Reddit's goal of being a "laissez-faire haven of (relatively) free expression"). Reddit has long positioned itself as a "free speech site with very few exceptions." hueypriest, Comment to *What Do You Think About Subreddits Such As /r/jailbait and /r/picsofdeadkids?*, REDDIT (July 20, 2011), http://www.reddit.com/r/IAmA/comments/iuz8a/iama_reddit_general_manager_ama/c26ukq8 [http://perma.cc/MW2Z-E4MR] (statement by a general manager at Reddit). Reddit has maintained this position even when said speech was personally revolting to its operators. *See* Adrian Chen, *Reddit CEO Speaks Out on Violentacrez in Leaked Memo: 'We Stand for Free Speech*,' GAWKER (Oct. 16, 2012, 6:36 PM), http://gawker.com/5952349/reddit-ceo-speaks-out-on-violentacrez-in-leaked-memo-we-stand-for-free-speech [http://perma.cc/F4TU-BPJV].

20. Olivia Solon, *Facebook Won't Block Fake News Posts Because It Has No Incentive, Experts Say*, GUARDIAN (Nov. 15, 2015, 6:52 PM), http://www.theguardian.com/technology/2016/nov/15/facebook-fake-news-us-election-trump-clinton [http://perma.cc/M67X-N4MS].

21. Mark Zuckerberg, FACEBOOK (Jan. 9, 2015), http://www.facebook.com/zuck/posts/10101844454210771 [http://perma.cc/P59J-E484].

22. Nick Statt, *Zuckerberg Calls Facebook 'A Platform for All Ideas' After Meeting with Conservatives*, VERGE (May 18, 2016, 7:39 PM), http://www.theverge.com/2016/5/18/11706266/mark-zuckerberg-facebook-conservative-news-censorship-meeting [http://perma.cc/EJA3-RCZM].

reference to their community or through user guidelines that prohibit certain undesirable, but not illegal, behavior.[23] For example, Reddit, which otherwise claims a laissez-faire approach to moderation, collaborates on moderation with a number of specific communities—including r/The_Donald, a subcommunity that vehemently and virulently supports the forty-fifth President of the United States.[24] YouTube prohibits the posting of pornography; at Facebook, community standards ban the posting of content that promotes self-injury or suicide.[25] None of their content policies stem from altruism. If users dislike the culture of a platform, they will leave and the platform will lose. For exactly that reason, platforms have taken measures—of varying efficacy—to police spam and harassment,[26] and in doing so to build a culture most amenable to mass engagement.

Ultimately, ambiguity serves neither the platforms nor their users. To users, hazy platform philosophy obscures any meaningful understanding of how platforms decide what is acceptable. Many wondered, in the wake of a recent leak, why Facebook's elaborate internal content moderation rules could justify deleting hate speech against white men, but allowed hate speech against black

---

23. *See, e.g.*, *Community Standards*, FACEBOOK, http://www.facebook.com/communitystandards [http://perma.cc/X9TP-PQZ5]; *Reddit Content Policy*, REDDIT, http://www.reddit.com/help/contentpolicy [http://perma.cc/H33F-Z339]; Snapchat Support, *Community Guidelines*, SNAPCHAT, http://support.snapchat.com/en-US/a/guidelines [http://perma.cc/DD39-LY4Q]; Twitter Help Ctr., *The Twitter Rules*, TWITTER, http://support.twitter.com/articles/18311 [http://perma.cc/KZD7-B8BK]; *Community Guidelines*, YOUTUBE, http://www.youtube.com/yt/policyandsafety/communityguidelines.html [http://perma.cc/F2BT-RPVQ].

24. *See* Menegus, *supra* note 19.

25. *See* sources cited *supra* note 23.

26. Twitter has banned a number of prominent "alt-right" accounts, for example, and has announced a set of tools to deal with harassment on the platform. Caitlin Dewey, *Twitter Has a Really Good Anti-Harassment Tool—And It's Finally Available to Everyone*, WASH. POST (Aug. 18, 2016), http://www.washingtonpost.com/news/the-intersect/wp/2016/08/18/twitter-has-a-really-good-anti-harassment-tool-and-its-finally-available-to-everyone [http://perma.cc/9RTA-8GE8]; Vijaya Gadde, *Twitter Executive: Here's How We're Trying To Stop Abuse While Preserving Free Speech*, WASH. POST (Apr. 16, 2015), http://www.washingtonpost.com/posteverything/wp/2015/04/16/twitter-executive-heres-how-were-trying-to-stop-abuse-while-preserving-free-speech [http://perma.cc/4U32-P35T]. *But see* Charlie Warzel, *Twitter Touts Progress Combating Abuse, But Repeat Victims Remain Frustrated*, BUZZFEED NEWS (July 20, 2017, 9:01 AM), http://www.buzzfeed.com/charliewarzel/twitter-touts-progress-combatting-abuse-but-repeat-victims [http://perma.cc/2TXQ-9XJU]; Charlie Warzel, *Twitter Is Still Dismissing Harassment Reports and Frustrating Victims*, BUZZFEED NEWS (July 18, 2017, 4:39 PM), http://www.buzzfeed.com/charliewarzel/twitter-is-still-dismissing-harassment-reports-and [http://perma.cc/4SHJ-MMYR].

children to remain online.[27] To platforms, philosophical indeterminacy over speech theories means there are few guiding stars to help navigate high-profile and rapidly evolving problems like fake news.

Moreover, even taking existing First Amendment theories separately, the fake news phenomenon illustrates how each theory fails to account for conspicuous phenomena that affect online speech. The marketplace theory, for example, fails to account for how easily accessible speech from many actors might change the central presumption that ideas compete to become true; the autonomy theory ignores that individuals are both speakers *and* listeners online; and the republican theory, in focusing only on political exchanges, casts aside much of the internet. All fail to account for how speech flows at a global and systemic scale, possibly because such an exercise would have been arduous if not impossible before social media platforms turned ephemeral words into indexed data.

These previously ephemeral interactions are now accessible to a degree of granularity that can enable new theories about how speech works globally at the systemic level. What insights might emerge if we focused on system-level operation, looking at the system from a descriptive standpoint? In the next Section, I will identify several systemic features of online speech, with a particular focus on how they are manipulated to produce fake news.

II. WHAT DOES THE SYSTEM TELL US ABOUT FAKE NEWS?

As the notable First Amendment and internet scholar Jack Balkin cautioned in 2004, "in studying the Internet, to ask 'What is genuinely new here?' is to ask the wrong question."[28] What matters instead is how digital technologies change the social conditions in which people speak, and whether that makes more salient what has already been present to some degree.[29] By focusing on what online platforms make uniquely relevant, we can discern social conditions that influence online speech, both desirable and not.

Below, I offer five newly conspicuous features that shape the ecosystem of speech online. Each of these features' manipulation exacerbates the fake news problem, but importantly none are visible—or addressable—under the marketplace, autonomy, or collectivist views of the First Amendment.

---

27. Julia Angwin & Hannes Grassegger, *Facebook's Secret Censorship Rules Protect White Men from Hate Speech but Not Black Children*, PROPUBLICA (June 28, 2017, 5:00 AM), http://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms [http://perma.cc/P9NV-A4SZ].

28. Balkin, *supra* note 5, at 2.

29. *Id.* at 2-3.

A. Filters

An obvious feature of online speech is that there is far too much of it to consume. Letting a thousand flowers bloom has an unexpected consequence: the necessity of information filters.[30]

The networked, searchable nature of the internet yields two interrelated types of filters. The first is what one might call a "manual filter," or an explicit filter, like search terms or Twitter hashtags. These can prompt misinformation: for example, if one searches "Obama birthplace," one will receive very different results than if one searches "Obama birth certificate fake." Manual filters can also include humans who curate what is accessible on social media, like content moderators.[31]

Less visible are implicit filters, for example algorithms that either watch your movements automatically or change based on how you manually filter. Such filters explain how platforms decide what content to serve an individual user, with an eye towards maximizing that user's attention to the platform. Ev Williams, co-founder of Twitter, describes this process as follows: if you glance at a car crash, the internet interprets your glancing as a desire for car crashes and attempts to accordingly supply car crashes to you in the future.[32] Engaging with a fake article about Hillary Clinton's health, for example, will supply more such content to your feed through the algorithmic filter.

That suggested content, sourced through the implicit filter, might also become more extreme. Clicking on the Facebook page designated for the Republican National Convention, as BuzzFeed reporter Ryan Broderick learned, led the "Suggested Pages" feature to recommend white power memes, a Vladimir Putin fan page, and an article from a neo-Nazi website.[33] It is this algorithmic

---

30. While there has always been competition for attention, in some fashion, the lowered cost of distribution and the abundance of information to be distributed alters the stakes of the competition. *See* J.M. Balkin, *Media Filters, the V-Chip, and the Foundations of Broadcast Regulation*, 45 DUKE L.J. 1131, 1132 (1996) ("In the Information Age, the informational filter, not information itself, is king.").

31. Adrian Chen, *The Human Toll of Protecting the Internet from the Worst of Humanity*, NEW YORKER (Jan. 28, 2017), http://www.newyorker.com/tech/elements/the-human-toll-of-protecting-the-internet-from-the-worst-of-humanity [http://perma.cc/AAK8-ZHEV].

32. David Streitfeld, *'The Internet Is Broken': @ev Is Trying To Salvage It*, N.Y. TIMES (May 20, 2017), http://www.nytimes.com/2017/05/20/technology/evan-williams-medium-twitter-internet.html [http://perma.cc/EJ4K-AEDC].

33. Ryan Broderick, *I Made a Facebook Profile, Started Liking Right-Wing Pages, and Radicalized My News Feed in Four Days*, BUZZFEED NEWS (Mar. 8, 2017), http://www.buzzfeed.com/ryanhatesthis/i-made-a-facebook-profile-started-liking-right-wing-pages-an [http://perma.cc/F5HQ-WGVJ].

pulling to the poles, rooted in a benign effort to keep users engaged, that unearths fake news otherwise relegated to the fringe.

B. Communities

Information filters, like the ones described above, have always existed in some form. We have always needed help in making sense of vast amounts of information. Before there were algorithms or hashtags, there were communities: office break rooms, schools, religious institutions, and media organizations are all types of community filters. The internet has changed, however, how digital communities can easily transcend the barriers of physical geography. The internet is organized in part by communities of interest, and information can thus be consumed within *and* produced by communities of distant but like-minded members. Both sides of this coin matter, especially for fake news.

Those focused on information consumption have long observed that filters can feed insular "echo chambers," further reinforced by algorithmic filtering.[34] Even if you are the only person you personally know who believes that President Barack Obama was secretly Kenyan-born, you can easily find like-minded friends online.

Notably, individuals also easily *produce* information, shared in online communities built around affinity, political ideology, hobbies, and more. At its best, this capability helps to remedy the historic shortcomings of traditional media: as danah boyd points out, traditional media outlets often do not cover stories like the protests in Ferguson, Missouri, in 2014, the Dakota Access Pipeline protests, or the disappearance of young black women until far too late.[35] At its worst, the capability to produce one's own news can cultivate a distrust of vaccines or nurture rumors about a president's true birthplace. Through developing their own narratives, these communities create their own methods to pro-

---

34. *See* Cass R. Sunstein & Adrian Vermeule, *Conspiracy Theories: Causes and Cures*, 17 J. POL. PHIL. 202 (2009); *see also* Daniel J. Isenberg, *Group Polarization: A Critical Review and Meta-Analysis*, 50 J. PERSONALITY & SOC. PSYCHOL. 1141 (1986) (explaining the interaction between group polarization and other social psychological phenomena). *But see* Richard Fletcher & Rasmus Kleis Nielsen, *Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication*, 67 J. COMM. 476 (2017) (finding no support for the idea that online audiences are more fragmented than offline audiences).

35. danah boyd, *Did Media Literacy Backfire?*, POINTS (Jan. 5, 2017), http://points.datasociety.net/did-media-literacy-backfire-7418c084d88d [http://perma.cc/K77L-EK3T].

duce, arrange, discount, or ignore new facts.[36] So, even though a television anchor might present you with a visual of Obama's American birth certificate, your online community—composed of members you trust—can present to you alternative and potentially more persuasive perspectives on that certificate.[37]

Taken together, this creates a bottom-up dynamic for developing trust, rather than focusing trust in top-down, traditional institutions.[38] In turn, that allows communities to make their own cloistered and potentially questionable decisions about how to determine truth—an ideal environment to normalize and reinforce false beliefs.

C. Amplification

The amplification principle explains how misinformation cycles through filters and permeates communities, which are in turn powered by the cheap, ubiquitous, and anonymous power of the internet. Amplification happens in two stages: first, when fringe ideas percolate in remote corners of the internet, and second, when those ideas seep into mainstream media.

Take, for example, the story of Seth Rich, a Democratic National Committee staffer found tragically murdered as a result of what Washington, D.C. police maintain was a botched robbery gone awry. WikiLeaks alluded to a connection between his unfortunate demise and his possibly leaking to them with little fanfare.[39] Weeks later, however, a local television affiliate in D.C. reported that a private investigator was looking into whether the murder was related to Rich allegedly providing email hacks of the Democratic National Committee to

---

36. Joshua Green, *No One Cares About Russia in the World Breitbart Made*, N.Y. TIMES (July 15, 2017), http://nytimes.com/2017/07/15/opinion/sunday/no-one-cares-about-russia-in-the-world-breitbart-made.html [http://perma.cc/YAX2-FJCQ].

37. *See* Tanya Chen & Charlie Warzel, *Here's How The Pro-Trump Media Is Handling The Don Jr. Emails*, BUZZFEED NEWS (July 11, 2017), http://www.buzzfeed.com/tanyachen/but-his-emails [http://perma.cc/A67B-XVRP]; Tarini Parti et al., *The Stories On Don Jr.'s Russia Meeting Are A "Bat Signal" For Trump's Base*, BUZZFEED NEWS (July 10, 2017), http://www.buzzfeed.com/tariniparti/the-stories-on-don-jrs-russia-meeting-were-a-bat-signal-for [http://perma.cc/D3DP-55RK].

38. *See* Charlie Warzel, *The Right is Building a New Media "Upside Down" to Tell Trump's Story*, BUZZFEED NEWS (January 23, 2017, 8:15 PM), http://www.buzzfeed.com/charliewarzel/the-right-is-building-a-new-media-upside-down-to-tell-donald [http://perma.cc/5VPT-HJFP].

39. Jeff Guo, *The Bonkers Seth Rich Conspiracy Theory, Explained*, VOX (May 24, 2017, 2:10 PM ET), http://www.vox.com/policy-and-politics/2017/5/24/15685560/seth-rich-conspiracy-theory-explained-fox-news-hannity [http://perma.cc/4CS5-JK3Q]; *see* Chris Cillizza, *The Tragic Death and Horrible Politicization of Seth Rich*, CNNPOLITICS (August 2, 2017, 1:43 PM ET), www.cnn.com/2017/08/02/politics/seth-rich-death-fox-news-trump [http://perma.cc/GJ6K-CJ97].

WikiLeaks.[40] Message boards on 4chan, 8chan, and Reddit grasped at these straws, launching their own vigilante investigations and further inquiries.[41] This is the first stage of amplification.

The second stage begins when those with a louder bullhorn observe the sheer volume of discussion, and the topic—true or not—becomes newsworthy in its own right. In the case of Rich, this happened when a number of prominent and well-networked individuals on Twitter circulated the conspiracy to their hundreds of thousands of followers using the hashtag #SethRich. That drew the attention of Fox News and its pundits, whose followers range in the millions, and in turn *Breitbart* and *Drudge Report*, which seed hundreds of blogs and outlets.[42]

The amplification dynamic matters for fake news in two ways. First, it reveals how online information filters are particularly prone to manipulation—for example, by getting a hashtag to trend on Twitter, or by seeding posts on message boards—through engineering the perception that a particular story is worth amplifying. Second, the two-tier amplification dynamic uniquely fuels perceptions of what is true and what is false. Psychologists tell us that listeners perceive information not only logically, but through a number of "peripheral cues" which signal whether information should be trusted. Cues can include whether the speaker is reliable (why trust in the source of information matters),[43] a listener's prior beliefs (why one's chosen communities matter),[44] and, most notably, the familiarity of a given proposition (why one's information

---

40. *See* sources cited *supra* note 39.

41. *See* sources cited *supra* note 39.

42. Charlie Warzel, *How One Pro-Trump Site Keeps a Debunked Conspiracy Theory Alive*, BUZZFEED NEWS (May 22, 2017), http://www.buzzfeed.com/charliewarzel/how-one-pro-trump-site-feeds-its-own-conspiracy-theories [http://perma.cc/EFX6-HLFU].

43. The classic study is ROBERT K. MERTON ET AL., MASS PERSUASION: THE SOCIAL PSYCHOLOGY OF A WAR BOND DRIVE (1946); see also RICHARD E. PETTY & JOHN T. CACIOPPO, ATTITUDES AND PERSUASION: CLASSIC AND CONTEMPORARY APPROACHES (1996) (identifying major approaches to attitude and belief change).

44. For important analysis of the factors that influence which ideas are accepted and which are not, see generally CHIP HEATH & DAN HEATH, MADE TO STICK: WHY SOME IDEAS SURVIVE AND OTHERS DIE (2008) (discussing how recipient understanding and memory of ideas are improved when such ideas are conveyed according to six factors); and Dan M. Kahan & Donald Braman, *Cultural Cognition and Public Policy*, 24 YALE L. & POL'Y REV. 149, 149-60 (2006) (arguing that multiple cultural factors strongly influence one's acceptance of ideas); *see also* Jared Wadley, *New Study Analyzes Why People Are Resistant to Correcting Misinformation, Offers Solutions*, UNIVERSITY OF MICHIGAN—MICHIGAN NEWS (Sept. 20, 2012), http://home.isr.umich.edu/sampler/new-study-analyzes-resistance-to-correcting-misinformation [http://perma.cc/54LG-7XKF] (examining the factors that allow the perpetuation of misinformation).

sources matter).[45] The latter point is crucial here: individuals are more likely to view repeated statements as true. (Advertising subsists on this premise: of course you will purchase the detergent you have seen before.)

Imagine, then, how many times a listener might absorb tidbits of the Seth Rich story: on talk radio on the way to work, through water cooler chat with a Reddit-obsessed co-worker, scrolling through Facebook, a scan of one's blogs, a group text, pundit shows promoting the conspiracy, or on the local television's evening news debunking it. Manifesting on that many platforms will, psychological research informs us, command attention and persuade. Even when something is as demonstrably bankrupt as the Seth Rich conspiracy, the false headline will be rated as more accurate than unfamiliar but truthful news.[46]

D. Speed

The staggering pace of sharing, and how it influences amplification, is particularly critical for understanding the spread of fake news.

Platforms are designed for fast, frictionless sharing. This function accelerates the amplification cycle explained above, but also targets it for maximum persuasion at each step. For example—before it was effectively obliterated from the internet[47]—a popular neo-Nazi blog called *The Daily Stormer* hosted a weekly "Mimetic Monday,"[48] where users posted dozens of image macros—the

---

45. The dominant account of this "illusory truth effect" is that familiarity increases the ease with which statements are processed (*i.e.*, processing fluency), which in turn is used heuristically to infer accuracy. Lynn Hasher et al., *Frequency and the Conference of Referential Validity*, 16 J. VERBAL LEARNING & VERBAL BEHAV. 107 (1977); *see also* Ian Maynard Begg et al., *Dissociation of Processes in Belief: Source Recollection, Statement Familiarity, and the Illusion of Truth*, 121 J. EXP. PSYCHOL.: GEN. 446 (1992) (reporting on experiments that concern the effect of repetition on the perceived truth of a statement).

46. *See, e.g.*, Gordon Pennycook et al., Prior Exposure Increases Perceived Accuracy of Fake News (July 6, 2017) (unpublished manuscript, Yale University), http://papers.ssrn.com/abstract_id=2958246 [http://perma.cc/T2TB-4F5P].

47. Keith Collins, *A Running List of Websites and Apps That Have Banned, Blocked, and Otherwise Dropped White Supremacists*, QUARTZ (Aug. 16, 2017), http://qz.com/1055141/what-websites-and-apps-have-banned-neo-nazis-and-white-supremacists [http://perma.cc/69TS-KKEK].

48. Alice Marwick and Rebecca Lewis, *Media Manipulation and Disinformation Online*, DATA&SOCIETY (2017) http://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf [http://perma.cc/268T-DE7H] [hereinafter Marwick & Lewis, *Media Manipulation*]; Alice Marwick & Rebecca Lewis, *The Online Radicalization We're Not Talking About*, SELECT/ALL (May 18, 2017, 11:16 AM), http://nymag.com/selectall/2017/05/the-online-radicalization-were-not-talking-about.html [http://perma.cc/QF7V-QV7R].

basis of memes—to be shared on Facebook, Twitter, Reddit, and other platforms.[49] Witty and eye-catching, if frequently appalling, macros like these allow rapid experimentation with talking points and planting ideas. Such efforts were responsible for spreading misinformation about French President Emmanuel Macron before the 2017 election.[50] This experimental factory is called "shitposting,"[51] and the fast, frictionless sharing across platforms is the machinery that helps the factory distribute at scale. Before social media platforms, this type of experimentation would have been phenomenally slow, or required resource-intensive focus groups.

Memes are a convenient way to package this information for distribution: they are easily digestible, nuance-free, scroll-friendly, and replete with community-reinforcing inside jokes. Automation software known as "bots"—whether directed by governments[52] or by people like the "Trumpbot overlord" named "Microchip"—are also often credited with circulating misinformation, because of how well they can trick algorithmic filters by exaggerating a story's importance.[53]

---

49. *See* Marwick & Lewis, *Media Manipulation*, *supra* note 48, at 36-37; *see also* Ryan Milner & Whitney Phillips, *A Meme Can Become a Hate Symbol by Social Consensus*, N.Y. TIMES (Oct. 3, 2016, 3:27 AM), http://www.nytimes.com/roomfordebate/2016/10/03/can-a-meme-be-a-hate-symbol-6/a-meme-can-become-a-hate-symbol-by-social-consensus [http://perma.cc/RR6H-3DG6] (discussing the use of the meme "Pepe the Frog" as an alt-right symbol); Christopher Paul & Miriam Matthews, *The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options To Counter It*, INT'L SECURITY & DEF. POL'Y CTR. (2016), http://www.rand.org/pubs/perspectives/PE198.html [http://perma.cc/A7G4-4GQK].

50. Ryan Broderick, *Here's How Far-Right Trolls Are Spreading Hoaxes About French Presidential Candidate Emmanuel Macron*, BUZZFEED NEWS (Apr. 25, 2017, 6:53 AM), http://www.buzzfeed.com/ryanhatesthis/heres-how-far-right-trolls-are-spreading-hoaxes-about [http://perma.cc/72Y9-UMKP]; Ryan Broderick, *Trump Supporters Online Are Pretending To Be French To Manipulate France's Election*, BUZZFEED NEWS (Jan. 24, 2017, 2:00 AM), http://www.buzzfeed.com/ryanhatesthis/inside-the-private-chat-rooms-trump-supporters-are-using-to [http://perma.cc/V7GX-UGJS].

51. Jason Koebler, *The Secret Chatrooms Where Donald Trump Memes Are Born*, MOTHERBOARD (Apr. 4, 2017, 12:54 PM), http://motherboard.vice.com/en_us/article/qk994b/the-secret-chatrooms-where-donald-trump-memes-are-born [http://perma.cc/ZY2C-KYVR]; Whitney Phillips et al., *Trolling Scholars Debunk the Idea That the Alt-Right's Shitposters Have Magic Powers*, MOTHERBOARD (Mar. 22, 2017, 11:56 AM), http://motherboard.vice.com/en_us/article/z4k549/trolling-scholars-debunk-the-idea-that-the-alt-rights-trolls-have-magic-powers [http://perma.cc/5F8F-ME8T].

52. *See* Bradshaw & Howard, *supra* note 11.

53. Joseph Bernstein, *Never Mind The Russians, Meet The Bot King Who Helps Trump Win Twitter*, BUZZFEED NEWS (Apr. 5, 2017, 2:01 AM), http://www.buzzfeed.com/josephbernstein/from-utah-with-love [http://perma.cc/FT7N-AAEG]; *see also* Robyn Caplan & danah boyd, *Mediation, Automation, Power*, DATA&SOCIETY (May 15, 2016), http://datasociety.net/pubs/ap/MediationAutomationPower_2016.pdf [http://perma.cc/DLF3-XA9E] (discuss-

Bots, however, are not the only ones to blame for rapid distribution. Almost sixty percent of readers share links on social media without even reading the underlying content.[54] Sharing on platforms is not only an exercise in communicating rational thought, but also signaling ideological and emotional affinity.

This explains, in part, why responses debunking fake news do not travel as quickly. For example, if one clicks on a story because one is already ideologically inclined to believe in it, there is less interest in the debunking—which likely means that the debunking would not even surface on one's feed in the first instance. It also explains why certain false ideas are so persistent: they are designed, in an effective and real-time laboratory, to be precisely that way.

E. Profit Incentives

Social media platforms make fake news uniquely lucrative. Advertising exchanges compensate on the basis of clicks for any article, which creates the incentive to generate as much content as possible with as little effort as possible. Fake news, sensational and wholly fabricated, fits these straightforward economic incentives. This yields everything from Macedonian teenagers concoct-

---

ing the power of social media outlet algorithms in "nudging" voters); Marc Fisher et al., *Pizzagate: From Rumor, to Hashtag, to Gunfire in D.C.*, WASH. POST (Dec. 6, 2016), http://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-indc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html [http://perma.cc/GND7-EXUF] (detailing the spread of the "Pizzagate" misinformation campaign); Philip Howard et al., *Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?*, OXFORD INTERNET INST. (Mar. 26, 2017), http://comprop.oii.ox.ac.uk/2017/03/26/junk-news-and-bots-during-the-u-s-election-what-were-michigan-voters-sharing-over-twitter [http://perma.cc/C7KH-8URD] (discussing the ability of "computational propaganda" to distribute large amounts of misinformation over social media platforms); Philip N. Howard & Bence Kollanyi, *Bots, #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum*, CORNELL UNIV. LIB. (June 20, 2016), http://arxiv.org/abs/1606.06356 [http://perma.cc/SA55-YFYY] (discussing the use of Twitter bots); Jared Keller, *When Campaigns Manipulate Social Media*, ATLANTIC (Nov. 10, 2010), http://www.theatlantic.com/politics/archive/2010/11/when-campaigns-manipulate-social-media/66351 [http://perma.cc/752N-QVWW] (detailing how political campaigns can use social media to trick algorithmic filters on search engines); J.M. Porup, *How Mexican Twitter Bots Shut Down Dissent*, MOTHERBOARD (Aug. 24, 2015, 7:00 AM), http://motherboard.vice.com/en_us/article/how-mexican-twitter-bots-shut-down-dissent [http://perma.cc/Z9ZY-27ME] (reporting on the use of twitter bots to attack government critics).

54. Caitlin Dewey, *6 in 10 of You Will Share This Link Without Reading It, a New, Depressing Study Says*, WASHINGTON POST (June 16, 2017), http://www.washingtonpost.com/news/the-intersect/wp/2016/06/16/six-in-10-of-you-will-share-this-link-without-reading-it-according-to-a-new-and-depressing-study [http://perma.cc/X26Z-W6VQ].

ing stories about the American election[55] to user-generated make-your-own-fake-news generators falsely claiming that specific Indian restaurants in London had been caught selling human meat.[56] These types of websites, particularly those that are hyperpartisan and thus primed to attract attention, have exploded in popularity: A BuzzFeed News study illustrated that over one hundred new pro-Trump digital outlets were created in 2016.[57]

There are two noteworthy elements to this uptick. First, the mechanics of advertising on these platforms facilitates the distribution of fake content: there is no need for a printing press, delivery trucks, or access to airtime. Cheap distribution means more money, only strengthening the incentive. Second, platforms render the appearance of advertisements and actual news almost identical.[58] This further muddies the water between what is financially motivated and what is not.

III. TOWARD A MORE ROBUST THEORY

Thinking in terms of the full system of speech—that is, considering filters, communities, amplification, speed, and profit incentives—gives us a far more detailed portrait of how misinformation flourishes online. It also provides a blueprint for what platforms are doing to curb fake news, all of which would make little sense under the more traditional theories described in Part I.

For example, platforms have exercised their ubiquitous filtering capabilities to target fake news. Google recently retooled its search engine to try to prevent

55. Craig Silverman & Lawrence Alexander, *How Teens in the Balkans Are Duping Trump Supporters with Fake News*, BUZZFEED NEWS (Nov. 3, 2016, 7:02 PM), http://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo [http://perma.cc/66QN-YYTR].

56. Craig Silverman & Sara Spary, *Trolls Are Targeting Indian Restaurants with a Create-Your-Own Fake News Site*, BUZZFEED NEWS (May 29, 2017, 2:58 PM), http://www.buzzfeed.com/craigsilverman/create-your-own-fake-news-sites-are-booming-on-facebook-and [http://perma.cc/6K7Z-E2PG].

57. Silverman & Alexander, *supra* note 55; *see also* Terrance McCoy, *Inside a Long Beach Web Operation That Makes up Stories about Trump and Clinton: What They Do for Clicks and Cash*, L.A. TIMES (Nov. 22, 2016, 1:30 PM), http://www.latimes.com/business/technology/la-fi-tn-fake-news-20161122-story.html [http://perma.cc/ZLT7-RJ7Y] (detailing a U.S. generator of "fake news").

58. Craig Silverman et al., *In Spite of the Crackdown, Fake News Publishers Are Still Earning Money from Major Ad Networks*, BUZZFEED NEWS (April 4, 2017, 9:05 AM), http://www.buzzfeed.com/craigsilverman/fake-news-real-ads [http://perma.cc/6G38-885M] (noting that "content-recommendation ad units, which provide ads made to look like real news headlines, were by far the most common ad format on the sites reviewed").

conspiracy and hoax sites from appearing in its top results,[59] while YouTube decided that flagged videos that contain controversial religious or supremacist content will be put in a limited state where they cannot be suggested to other users, recommended, monetized, or given comments or likes.[60] And Facebook has partnered with fact-checkers to flag conspiracies, hoaxes, and fake news; flagged articles are less likely to surface on users' news feeds.[61] These tweaks, at least conceptually, should influence the algorithmic filters that yield information.

Similarly, Facebook has overtly recognized that speed and amplification can contribute to misinformation. It now deprioritizes links that are aggressively shared by suspected spammers, on the theory that these links "tend to include low quality content such as clickbait, sensationalism, and misinformation."[62] Facebook is also launching features that push users to think twice before sharing a story, by juxtaposing their link with other selected "Related Articles."[63] Twitter specifically targets bots, looking for those that may game its system to artificially raise the profile of misinformation and conspiracies.[64]

Recognizing the profit element, Google and Facebook have both barred fake news websites from using their respective advertising programs.[65] Face-

---

59. Nicas, *supra* note 65; *see also* Peter Lloyd, *Google Introduces New Global Fact-Checking Tag To Help Filter 'Fake News,'* DAILY MAIL (Apr. 7, 2017, 6:05 AM), http://www.dailymail.co.uk/sciencetech/article-4389436/Google-introduces-new-global-fact-checking-tags.html [http://perma.cc/8PMT-D8MB] (detailing Google's program to inform readers about the credibility of sources).

60. The YouTube Team, *An Update on Our Commitment To Fight Terror Content Online*, YOUTUBE (August 1, 2017), http://youtube.googleblog.com/2017/08/an-update-on-our-commitment-to-fight.html [http://perma.cc/S3PM-23LK].

61. Laura Hazard Owen, *Clamping Down on Viral Fake News, Facebook Partners with Sites like Snopes and Adds New User Reporting*, NIEMANLAB (Dec. 15, 2016), http://www.niemanlab.org/2016/12/clamping-down-on-viral-fake-news-facebook-partners-with-sites-like-snopes-and-adds-new-user-reporting [http://perma.cc/U2XC-JB5A].

62. Adam Mosseri, *News Feed FYI: Showing More Informative Links in News Feed*, FACEBOOK NEWSROOM (June 30, 2017), http://newsroom.fb.com/news/2017/06/news-feed-fyi-showing-more-informative-links-in-news-feed [http://perma.cc/L73X-K6QX].

63. Kaya Yurieff, *Facebook Steps Up Fake News Fight with "Related Articles,"* CNN (Aug. 3, 2017, 2:27 PM ET), http://money.cnn.com/2017/08/03/technology/facebook-related-articles [http://perma.cc/LTM9-24VW].

64. Nicholas Thompson, *Instagram Unleashes an AI System to Blast Away Nasty Comments*, WIRED (June 29, 2017, 9:00 AM), http://www.wired.com/story/instagram-launches-ai-system-to-blast-nasty-comments [http://perma.cc/D8HV-BAHF].

65. After the 2016 election, Google began barring fake news websites from its AdSense advertising program. Jack Nicas, *Google to Bar Fake-News Websites from Using Its Ad-Selling Software*, WALL ST. J, (Nov. 14, 2016, 9:55 PM), http://www.wsj.com/articles/google-to-bar-fake-news-websites-from-using-its-ad-selling-software-1479164646 [http://perma.cc/PE5R-WFXG]. Facebook soon followed suit. Deepa Seetharaman, *Facebook Bans Fake News Sites*

book has also eliminated the ability to spoof domains pretending to be real publications to profit from those who click through to the underlying sites, which are replete with ads.[66] This may speak to profit-oriented fake news, but not to propaganda and misinformation that is fueled by nonfinancial incentives.[67]

These systemic features can also help us interrogate concepts whose definitions have long been assumed. Take, for example, the concept of censorship. Traditionally, and in the speaker-focused marketplace and autonomy theories, censorship evokes something very specific: blocking the articulation of speech. As prominent sociologist Zeynep Tufekci argues, however, censorship now operates via information glut—that is, drowning out speech instead of stopping it at the outset.[68] As with the Saudi Arabian example referenced above, Tufekci points to the army of internet trolls deployed by the Chinese and Russian governments to distract from critical stories and to wear down dissenters through the manipulation of platforms.[69] If platforms are the epicenters of this new censorship, misinformation is the method: the point of censorship by disinformation is to destroy attention as a key resource.[70] What results, Tufekci explains, is a "frayed, incoherent, and polarized public sphere that can be hostile to dissent."[71] This all becomes visible when information filters are taken into account.

---

from Using Its Advertising Network, WALL ST. J. (Nov. 14, 2016, 9:14 PM), http://www.wsj.com/articles/facebook-bans-fake-news-sites-from-using-its-advertising-network-1479175778 [http://perma.cc/XYK3-PSKW].

66. Adam Mosseri, *News Feed FYI: Addressing Hoaxes and Fake News*, FACEBOOK NEWSROOM (Dec. 15, 2016), http://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news [http://perma.cc/S7JF-L2V6].

67. Mark Verstraete et al., *Identifying and Countering FAKE NEWS*, UNIV. OF ARIZ. SCH. OF L. (2017), http://law.arizona.edu/sites/default/files/asset/document/fakenewsfinal_0.pdf [http://perma.cc/L6VZ-AVFG].

68. Zeynep Tufekci, *WikiLeaks Isn't Whistleblowing*, N.Y. TIMES (Nov. 4, 2016), http://www.nytimes.com/2016/11/05/opinion/what-were-missing-while-we-obsess-over-john-podestas-email.html [http://perma.cc/5FER-M2WK].

69. Knight First Amendment Institute, *Free Speech in the Networked World*, YOUTUBE (May 4, 2017), http://www.youtube.com/watch?v=IhldzzehTcs; *see also* Bradshaw & Howard, *supra* note 11 (discussing the potential for governments to promote misinformation and manipulate public opinion through social media).

70. Attention, as a critical and scarce resource, already faces great demand from another industry: modern advertising. TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (2016).

71. ZEYNEP TUFEKCI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 240 (2017).

It would be easy to conclude that platforms—best positioned to address the aforementioned features—should alone shoulder the burden to prevent fake news. But asking private platforms to exercise unilateral, unchecked control to censor is precarious.[72] Few factors would constrain possible abuses. For example, Jonathan Zittrain raises the possibility of Facebook manipulating its end-users by using political affiliation to alter voting outcomes—something that could be impervious to liability as protected political speech.[73] No meaningful accountability mechanism exists for these platforms aside from public outcry, which relies on intermediaries to divine what platforms are actually doing. And yet, the other extreme—a content-neutral and hands-off approach—offers empty guidance in the face of organized fake news or other forms of manipulation.

Instead, we must collectively build a theory that accounts for these shifting sands, one that provides workable ideals rooted in reality. Scaffolding for that theory can be found in what Balkin has termed the "democratic culture" theory, which seeks to ensure that each individual can meaningfully participate in the production and distribution of culture.[74] A focus on culture, not politics, does more than remedy the central gap of the collectivist view while maintaining its system-wide focus. It also helps us expand our focus beyond legal theory to relevant disciplines like social psychology, sociology, anthropology, and cognitive science. For example, once we understand amplification as a relevant concept, we should account for the psychology of how people actually come to be-

---

72. *See* Mike Masnick, *Nazis, The Internet, Policing Content and Free Speech*, TECHDIRT (Aug. 25, 2017, 10:54 AM), http://www.techdirt.com/articles/20170825/01300738081/nazis-internet-policing-content-free-speech.shtml [http://perma.cc/8QLT-DT77]; *cf.* Matthew Prince, *Why We Terminated Daily Stormer*, CLOUDFLARE (Aug. 16, 2017), http://blog.cloudflare.com/why-we-terminated-daily-stormer [http://perma.cc/6BBA-D43S] (explaining the practical and deliberative considerations that went into terminating the Daily Stormer).

73. For example, Jonathan Zittrain raises the possibility of Facebook manipulating its end-users based on political affiliation to alter voting outcomes—something that could be impervious to liability as protected political speech. Jonathan Zittrain, *Response, Engineering an Election: Digital Gerrymandering Poses a Threat to Democracy*, 127 HARV. L. REV. F. 335, 335-36 (2014); Jonathan Zittrain, *Facebook Could Decide an Election Without Anyone Ever Finding Out*, NEW REPUBLIC (June 1, 2014), http://www.newrepublic.com/article/117878/information-fiduciary-solution-facebook-digital-gerrymandering [http://perma.cc/U479-WVPU]; *see also* Timothy Revell, *How To Turn Facebook into a Weaponised AI Propaganda Machine*, NEW SCIENTIST (July 28, 2017), http://www.newscientist.com/article/2142072-how-to-turn-facebook-into-a-weaponised-ai-propaganda-machine [http://perma.cc/VX7B-F7KG] (examining the potential of social media platforms like Facebook to disseminate propaganda and impact elections) .

74. Balkin, *supra* note 5 at 4; *see also id.* at 33 ("A 'democratic' culture, then, means much more than democracy as a form of self-governance. It means democracy as a form of social life in which unjust barriers of rank and privilege are dissolved, and in which ordinary people gain a greater say over the institutions and practices that shape them and their futures.").

lieve what is true—not only through rational deliberation, but also by using familiarity and in-group dynamics as a proxy for truth. Building on this frame will require more meaningful information from the platforms themselves.

A clear theory is more important now than ever. For one, a functioning theory can bridge the widening gap of expectations between what a platform permits and what the public expects. Practically, an overarching theory can also help navigate evolving social norms. Platforms make policy decisions based on contemporary norms: for example, until recently choosing to target and takedown accounts linked to foreign terrorists but not those linked to white nationalists and neo-Nazis, even though both types of organizations perpetuate fake news domestically.[75] We have to understand how definitions of tricky and dynamic concepts, like fake news, are created, culturally contingent, and capable of evolution. Finally, and crucially, we need a theory to help direct and hold accountable the automated systems that increasingly govern speech online. These systems will embed cultural norms into their design, and enforce them through implicit filters we cannot see. Only with a cohesive theory can we begin to resolve the central conundrum confronting social media platforms: they are private companies that have built vast systems sustaining the global, networked public square, which is the root of both their extraordinary value and their damnation.

Nabiha Syed is an assistant general counsel at BuzzFeed, a visiting fellow at Yale Law School, and a non-resident fellow at Stanford Law School. *All of my gratitude goes to Kate Klonick, Sabeel Rahman, Sushila Rao, Noorain Khan, Azmat Khan, Emily Graff, Alex Georgieff, Sara Yasin, Smitha Khorana, and the staff of the* Yale Law Journal Forum *for their incisive comments and endless patience, and to Nana Menya Ayensu, for everything always, but especially the coffee.*

Preferred Citation: Nabiha Syed, *Real Talk About Fake News: Towards a Better Theory for Platform Governance*, 127 YALE L.J. F. 337 (2017), http://www.yalelawjournal.org/forum/real-talk-about-fake-news.

75. J.M. Berger, *Nazis vs. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks*, GEO. WASH. PROGRAM ON EXTREMISM (Sept. 2016), http://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Nazis%20v.%20ISIS%20Final_0.pdf [http://perma.cc/7A3D-EBBP].