1-18-2018

Speech v. Speakers

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

Speech v. Speakers

Thomas Kadri

Slate's Future Tense – January 18, 2018

2017 was quite a year for Twitter. The platform gave everyone an extra 140 characters to play with in every tweet. It was accused of helping Russia interfere in our elections. It was even endorsed by the White House as President Trump's official mouthpiece.

But Twitter's most significant moments didn't prompt breaking-news alerts. In November, it began targeting far-right activists by revoking their "verified" badges—the little blue icons Twitter uses to say that "an account of public interest is *authentic*." Alt-right darlings like Jason Kessler lost their checkmarks because Twitter worried that the badge could be seen as an "endorsement" of their hateful views. So, from now on, users will be deverified for "promoting hate" or even for "supporting organizations or individuals" that promote hate.

Then, in December, Twitter began suspending renowned white supremacists and other far-right hatemongers. Indelicately dubbed the "#TwitterPurge" (the suspensions began on Joseph Stalin's birthday), Twitter said it hoped to "reduce hateful and abusive content" by chucking out racists and extremists. Most significantly, Twitter admitted that it will now consider conduct occurring "off the platform" in making suspension decisions—including whether a user affiliates with organizations that promote violence. And, in another important step, Twitter users may now report entire profiles, not merely offensive tweets.

These rule changes were widely celebrated. Calls to "ban the Nazis" had been growing, particularly after the horrors in Charlottesville, Virginia. Within Twitter itself, there was delight that the company was no longer the "free speech wing of the free speech party." "No more nazis," crowed one employee when the new verification policy was announced. The only real concern seemed to be that the new rules could create a "slippery slope" to banning left-leaning activists whose affiliations aren't always so pacifist.

The debate, then, has focused on *which* users to ban and deverify. When President Trump shared three anti-Muslim videos from a British ultranationalist group, some questioned why the group's leaders still had verified badges while their American

alt-right peers had lost theirs.
Similarly, some asked why infamous white nationalist Richard Spencer survived the "purge" when infamous white nationalist Jeff Schoep got the boot.

Although these are tempting questions to ask, the far more interesting issue here isn't which users Twitter should ban or deverify. We should instead focus on how Twitter's new rules blur the line between extremist *speakers* and their *speech*. This is a subtle but crucial shift in the way Twitter governs who can be a part of its platform and what can get you kicked off it.

Twitter's new approach has changed the rules of the game. The platform's content moderators, who already faced a daunting task in policing offensive tweets, must now make complex judgments about what people are doing offline, including whether they're associating with odious groups or individuals. And Twitter's users can now flag entire accounts at the click of a button. No longer does Twitter care only about harmful speech on its platform—it's now targeting hateful people and their hateful conduct. This shift puts the platform knee-deep in the quagmire of moral signaling, of telling the world whom the company favors and disfavors, of answering the public's calls to delegitimize bigots for their bigotry.

These might seem like valid goals for a social network. After all, Twitter is trying to create a community to which its users would like to belong. But the platform's new rules should concern anyone who cares about private governance of online speech. Twitter has set itself up as judge, jury, and executioner. A person can now be removed from the platform without ever spewing a single offensive tweet. Exactly how Twitter will manage this new system remains opaque. In December, a company spokesperson said it would rely on first-person and bystander reports of offline mischief, as well as on consultations with experts to identify accounts that merit permanent suspension. But the finer details of how Twitter will reliably investigate accusations of real-world transgressions haven't been released (if they exist).

Even if Twitter could design a robust and transparent system to make these decisions, dangers remain. This kind of peer-to-peer policing can pose problems for free speech. Social networks have, of course, regularly asked their users to flag inappropriate content. But when the upshot of reporting a tweet is that 280 characters disappear from the platform, the stakes are much lower. Under the new rules, though, entire accounts might be blocked and deverified based on public outcry about a user's off-platform actions or affiliations. In legal terms, we might call this a "heckler's veto"—the First Amendment idea that the government can't

silence a speaker merely because a large crowd complains about her speech. In other settings, we might call it vigilantism, where the public takes law enforcement into its own hands. Whatever we call it in the Twittersphere, it could result in rampant censorship if implemented imprecisely.

This, in turn, raises another constitutional concept: prior restraint. Under the First Amendment, courts will rarely muzzle a speaker before he has actually said anything. Instead, speakers may generally say their piece and face the consequences afterward, as, for example, could happen if President Trump sued Michael Wolff for defamation once *Fire and Fury* hit the bookshelves. Under Twitter's new rules, people could be blocked from ever tweeting again based not on what they said but on what their offline affiliations suggest they *might* say.

What's more, Twitter's newfangled verification rules will make it harder for the public to tell real accounts from fake ones. By relying on deverification to undermine the online legitimacy of controversial users, other users are robbed of a valuable tool. Twitter's blue badge used to serve a vital purpose: telling the public which accounts were authentic and which were not. But now that all sorts of conduct can result in deverification, the checkmark is no longer a trustworthy indicator of fakery.

These concerns aren't hypothetical. A few weeks ago, confusion erupted when several journalists—including USA Today's Supreme Court correspondent—reported that retired Supreme Court Justice David Souter tweeted criticism of another judge over a divisive decision. Adam Liptak of the New York Times was quick to label the account as "fake" (Justice Souter is a renowned technophobe), but that didn't stop the false story spreading on Twitter and beyond.

The purported Justice Souter account had no blue checkmark, so what's the big deal? Now that Twitter has muddled the question of verification with its judgments about viewpoint, the badge is sapped of reliability—and conversely, the *lack* of a badge doesn't necessarily signify inauthenticity. Twitter's justification for the rule change was that people might misconstrue verification as a validation of the user's tweets (even though the platform steadfastly maintains that the verified badge "does not imply an endorsement by Twitter"). But if Twitter truly fears people will stumble upon, say, Richard Spencer's posts and assume that the company, too, is "proudly White," then Twitter should create a new badge to signal its distaste for noxious user content. It shouldn't impede people's ability to distinguish the real accounts from the fake ones. In the age of bot armies and fake news, parsing truth

from fiction online has never been harder and yet more important. Twitter isn't helping.

Of course, the problems raised by Twitter's new rules aren't actually *constitutional* ones. There's been no action by the government, which is necessary for a First Amendment claim. (Tough luck, Chuck Johnson.) But that doesn't mean we can't learn valuable lessons from the principles underlying the Constitution's guarantee of free speech. With so much of our public debate happening on platforms like Twitter, it's essential that our private governors tread carefully when they decide who can participate in the "modern public square."