

Why Twitter labeling Trump's tweets as "potentially misleading" is a big step forward

From time to time a really bad post on a social network gets a lot of attention. Say a head of state falsely accuses a journalist of murder, or suggests that mail-in voting is illegal – those would be pretty bad posts, I think, and most people working inside and outside of the social network could probably agree on that. In my experience, though, average people and tech people tend to think very differently about what to do about a post like that. Today I want to talk about why.

When an average person sees a very bad post on a social network, they may call for it to be removed immediately. They will justify this removal on moral grounds – keeping the post up, they will say, is simply *indecent*. To leave it up would reflect poorly on the moral character of everyone who works at the company, especially its top executives. Some will say the executives should resign in disgrace, or possibly be arrested. Congress may begin writing letters, and new laws will be proposed, so that such a bad post never again appears on the internet.

When a tech company employee sees a really bad post, they are just as likely to be offended as the next person. And if they work on the company's policy team, or as a moderator, they will look to the company's terms of services. Has a rule been broken? Which one? Is it a clear-cut violation, or can the

post be viewed multiple ways?

If a post is deeply offensive but not covered by an existing rule, the company may write a new one. As it does, employees will try to write the rule narrowly, so as to rule in the maximum amount of speech, while ruling out only the worst. They will try to articulate the rule clearly, so that it can be understood in every language by an army of low-paid moderators. (And who may be developing post-traumatic stress syndrome and related conditions.)

Put another way, when an average person sees a really bad post, their instinct is to react with anger. And when a tech person sees a really bad post, their instinct is to react *practically*.

All of that context feels necessary to understand two Twitter debates playing out today: one over what Twitter ought to do about the fact that President Trump keeps tweeting without evidence that one of the few high-profile Republicans who regularly speaks out about him, the onetime congressman and current MSNBC host Joe Scarborough, may be implicated in the 2001 death of a former staffer. And one over what to do about the president's war on voting by absentee ballot.

As to the former: In fact, according to the medical examiner, former Scarborough aide Lori Klausutis died of a blood clot.) Now her widow is petitioning Twitter CEO Jack Dorsey to remove Trump's tweets suggesting there may have been foul play. John Wagener wrote up the day's events in the *Washington Post*:

With no evidence, Trump has continued to push a conspiracy theory that Scarborough, while a member of Congress, had an affair with his married staffer and that he may have killed her – a theory that has been debunked by news organizations including The Washington Post and that Timothy Klausutis called a “vicious lie” in his letter to Dorsey.

On Tuesday morning, Trump went on Twitter again to advocate

the “opening of a Cold Case against Psycho Joe Scarborough,” which he said was “not a Donald Trump original thought.”

“So many unanswered & obvious questions, but I won’t bring them up now!” Trump added. “Law enforcement eventually will?”

If you believe social networks are obligated to remove posts that are indecent, it’s clear why you would want these tweets to come down. The president is inflicting an emotional injury on an innocent, bereaved man for political gain. (Trump has historically benefitted from falsely suggesting his Republican opponents are murderers, as Jonathan Chait notes here.)

But if your job is to write or enforce policy at a tech company, your next steps are far less clear. Consider the facts. Did Trump say definitively that Scarborough committed murder? He didn’t – “maybe or maybe not,” he tweeted this morning. Did Trump incite violence against Scarborough, directly or indirectly? (Twitter has promised to hide such tweets behind a warning label, but it has never done so.) I don’t think so, and while encouraging law enforcement to investigate the case arguably represents an abuse of presidential power, our nation’s founders invested the responsibility for reining in a wayward chief executive not with private companies but with the other two branches of government.

Let’s make it more complex: Scarborough is a public figure – a former congressman, no less. Traditionally social networks have tolerated much more indecency when it comes to average people wanting to yell at the rich and powerful, and when it comes to the rich and powerful yelling at one another. And when two of those figures are engaged in political discourse – the kind of discourse that the First Amendment, which informs so many of the principles of tech company speech policies, sought to protect above all else – a tech policy person would probably want to give that speech the widest possible

latitude.

I spent the day talking with former Twitter employees who worked on speech and policy issues. For the most part, they thought Trump's Scarborough tweets should stay up. For one, the tweets don't violate existing policy. And two, they believe you can't design a policy that bans these tweets that doesn't also massively chill speech across the platform. As one former employee put it to me, "If speculation about unproven crime is not allowed, I have bad news for anyone who wants to tweet about a true crime podcast."

Now, it's possible for me to imagine a time when Twitter would have to take action against these tweets. There was a time when Alex Jones' tweets and videos about the Sandy Hook school shooting also fell into the realm of "speculating about true crime," even though his conspiracy theories were almost certainly promoted in bad faith. But then Jones' fans began stalking and harassing families of the murder victims, in some cases threatening to kill them. Eventually Jones was removed from most of the big social platforms.

If Trump continues to promote the lie about Scarborough, we can assume some of his followers will take matters into their own hands. It's been barely a year since one of those followers was sentenced to 20 years in prison for mailing 16 pipe bombs to people he perceived to be Trump's enemies. If something similar happens as a result of the Scarborough tweets, Twitter will face criticism for failing to act. It's a terrible position for the company to be in.

But mostly it's just a terrible thing for the president to *do*. And in a democracy we have remedies for bad behavior that go well beyond asking a tech company to de-platform a politician. You can speak your mind, you can march in the streets, and you can vote. That's why, for most problems of political speech, my preferred solution is more speech, in the form of more votes.

Which brings us to the day's surprising conclusion: Twitter's decision to label, for the first time, some of Trump's tweets as potentially misleading. Makena Kelly has the story in *The Verge*:

On Tuesday, Twitter labeled two tweets from President Donald Trump making false statements about mail-in voting as "potentially misleading." It's the first time the platform has fact-checked the president.

The label was imposed on two tweets Trump posted Tuesday morning falsely claiming that "mail-in ballots will be anything less than substantially fraudulent" and would result in "a rigged election." The tweets focused primarily on California's efforts to expand mail-in voting due to the novel coronavirus pandemic. On Sunday, the Republican National Committee sued California Gov. Gavin Newsom over the state's moves to expand mail-in voting.

According to a Twitter spokesperson, the tweets "contain potentially misleading information about voting processes and have been labeled to provide additional context around mail-in ballots." When a user sees the tweets from Trump, a link from Twitter is attached to them that says "Get the facts about mail-in ballots." The link leads to a collection of tweets and news articles debunking the president's statements.

This story is surprising for several reasons. It involves Twitter, a company notoriously prone to inaction, making a decisive move against its most powerful individual user. It ensures a long stretch of partisan mud-wrangling over which future tweets from which other politicians deserve similar treatment – and over whether one side or another is being punished disproportionately. And it puts Twitter prominently in the position it has long sought to avoid – "the arbiter of truth," chiming in when the president lies to say that no,

actually, it's legal to vote by absentee ballot.

And yet at the same time, Twitter's decision was rooted in principle. In January Twitter began allowing users to flag tweets that contain misleading information about how to vote. Today it applied that policy, fairly and with relative precision. Some have criticized the design and wording of the actual label – “Get the facts about mail-in ballots” doesn't exactly scream “the president is lying about this.” But it still feels like a step forward, and not a small one.

Social networks that reach global scale will always suffer from really bad posts, some of them posted by their most prominent users. And it's precisely because those platforms have become so important to political speech that I would rather decisions about what stays up and what comes down not be dictated by the whims of unelected, unaccountable founders.

Twitter's decision to leave up some of Trump's awful tweets and label others as misleading won't fully satisfy anyone. But in my view this is a case where the company has made some hard decisions in a relatively judicious way. And anyone who tries to write a better, more consistent policy – one that goes beyond “this is indecent, take it down” – will find that it's much harder than it looks.

The Ratio

Today in news that could affect public perception of the big tech platforms.

↑ **Trending up: Facebook** announced new features for Messenger that will alert users about messages that appear to come from financial scammers or child abusers. The company said the detection will occur only based on metadata—not analysis of the content of messages—so that it doesn't undermine end-to-end encryption. (Andy Greenberg / *Wired*)

↓ **Trending down:** YouTube deleted comments with two phrases that insult the Chinese Communist party. The company said it was an error. (James Vincent / *The Verge*)

↓ **Trending down:** Amazon supplied local TV news stations with a propaganda reel intended to change the subject from deaths and illnesses at its distribution centers. At least 11 stations aired it, and this video lets you watch various news anchors robotically parrot the PR talking points. (Nick Statt / *The Verge*)

Virus tracker

Total cases in the US: More than 1,685,800

Total deaths in the US: At least 98,800

Reported cases in California: 99,547

Total test results (positive and negative) in California:
1,696,396

Reported cases in New York: 368,669

Total test results (positive and negative) in New York:
1,774,128

Reported cases in New Jersey: 155,764

Total test results (positive and negative) in New Jersey:
635,892

Reported cases in Illinois: 113,402

Total test results (positive and negative) in Illinois:
786,794

Data from *The New York Times*. Test data from *The COVID Tracking Project*.

Governing

□ **Facebook** spent years studying how the platform polarized people, according to sources and internal documents. One slide from a 2018 presentation read "our algorithms exploit the human brain's attraction to divisiveness." Here are Jeff Horwitz and Deepa Seetharaman from the *Wall Street Journal*:

Facebook had kicked off an internal effort to understand how its platform shaped user behavior and how the company might address potential harms. Chief Executive Mark Zuckerberg had in public and private expressed concern about "sensationalism and polarization."

But in the end, Facebook's interest was fleeting. Mr. Zuckerberg and other senior executives largely shelved the basic research, according to previously unreported internal documents and people familiar with the effort, and weakened or blocked efforts to apply its conclusions to Facebook products.

Facebook policy chief Joel Kaplan, who played a central role in vetting proposed changes, argued at the time that efforts to make conversations on the platform more civil were "paternalistic," said people familiar with his comments.

President Trump is considering creating a panel to review complaints of anticonservative bias on social media. **Facebook**, **Twitter**, and **Google** all pushed back against the proposed panel, denying any anticonservative bias. I imagine today's action from Twitter will come up, if this thing turns out to be real. (John D. McKinnon and Alex Leary / *The Wall Street Journal*)

Doctors with verified accounts on **Facebook** are spreading coronavirus misinformation. The company has been trying to crack down on misinformation about virus, but the accounts are

still able to reach hundreds of thousands of people regularly.
(Rob Price / *Business Insider*)

Here's a guide to the most notorious spin doctors and conspiracy theorists spreading misinformation about the coronavirus pandemic. (Jane Lytvynenko, Ryan Broderick and Craig Silverman / *BuzzFeed*)

Influencers say **Instagram** is biased against plus-sized bodies – and they might be right. Content moderation on social media is usually a mix of artificial intelligence and human moderators, and both methods have a potential bias against larger bodies. (Lauren Strapagiel / *BuzzFeed*)

Joe Biden's digital team is trying to raise his online profile prior to the 2020 election while understanding his limitations on social media. Which is another way of saying he's still not on TikTok. (Sam Stein / *Daily Beast*)

Democrats are introducing a new bill that would tighten restrictions on online political ad-targeting on platforms like **Facebook**. The Protecting Democracy from Disinformation Act would limit political advertisers to targeting users based only on age, gender and location – a move intended to crack down on microtargeting. (Cristiano Lima / *Politico*)

Two new laws in Puerto Rico make it a crime to report information about emergencies that the government considers "fake news." The ACLU filed a lawsuit on behalf of two Puerto Rican journalists who fear the laws will be used to punish them for their reporting on the coronavirus crisis. (Sara Fischer / *Axios*)

One of the first contact-tracing apps in the US, North and South Dakota's Care19, violates its own privacy policy by sharing location data with an outside company. The oversight suggests that state officials and **Apple**, both of which were responsible for vetting the app before it became available April 7th, were asleep at the wheel. (Geoffrey A. Fowler / *The*

Washington Post)

China's virus-tracking apps have been collecting information, including location data, on people in hundreds of cities across the country. But the authorities have set few limits on how that data can be used. And now, officials in some places are loading their apps with new features, hoping the software will live on as more than just an emergency measure. (Raymond Zhong / *The New York Times*)

Serious security vulnerabilities were discovered in Qatar's mandatory contact tracing app. The security flaw, which has now been fixed, would have allowed bad actors to access highly sensitive personal information, including the name, national ID, health status and location data of more than one million users. (Amnesty International)

Inside the NSA's secret tool for mapping your social network. Edward Snowden revealed the agency's phone-record tracking program. But the database was much more powerful than anyone knew. (Barton Gellman / *Wired*)

Silicon Valley's main data-protection watchdog in Europe came under attack for taking too long to wrap up probes into **Facebook**, **Instagram** and **WhatsApp**. The group has yet to issue any significant fines two years after the EU empowered it to levy hefty penalties for privacy violations. (Stephanie Bodoni / Bloomberg)

A court in the Netherlands is forcing a grandmother to delete photos of her grandkids that she posted on Facebook and Pinterest without their parents' permission. The judge ruled the matter was within the scope of the EU's General Data Protection Regulation. (BBC)

Industry

□ Shopping for **Instacart** is dangerous during the pandemic. Now,

workers who've gotten sick say they haven't been able to get the quarantine pay they were promised. Russell Brandom at *The Verge* has the story:

It's a common story. On forums and in Facebook groups, Instacart's sick pay has become a kind of sour joke. There are lots of posts asking how to apply, but no one seems to think they'll actually get the money. The Verge spoke to eight different workers who were placed under quarantine – each one falling prey to a different technicality. A worker based in Buffalo was quarantined by doctors in March but didn't qualify for an official test, leaving him with no verification to send to reps. In western Illinois, a man received a quarantine order from the state health department, but without a test, he couldn't break through. Others simply fell through the cracks, too discouraged to fight the claim for the weeks it would likely take to break through.

Amazon lost some online shoppers to rivals during the pandemic as it struggled to keep up with demand. Now the retail giant is turning back to faster shipping times and big sales to lure people back to the platform. (Karen Weise / *The New York Times*)

Google said the majority of its employees will work from home through 2020. It's giving everyone \$1,000 to cover any new work-from-home expenses. (Chaim Gartenberg / *The Verge*)

Welcome to the age of the **TikTok** cult. These aren't the ideological cults most people are familiar with. Instead, they are open fandoms revolving around a single creator. Right now they're being weaponized to perform social-media pranks, but it feels like something much darker is around the corner. (Taylor Lorenz / *The New York Times*)

Zoom temporarily removed **Giphy** from its chat feature, days after **Facebook** acquired the GIF platform for \$300 million. "Once additional technical and security measures have been

deployed, we will re-enable the feature” the company said.

Facebook renamed Calibra, the digital wallet it hopes will one day be used to access the Libra digital currencies, to “Novi.” The company said that the new name was inspired by the Latin words “novus” and “via,” which mean “new” and “way” – and not, as I had assumed, the English words “non” and “viable.” (Jon Porter / *The Verge*)

Facebook’s internal R&D group launched a new app called CatchUp that makes it easier for friends and family in the US to coordinate phone calls with up to 8 people. I do not get this one at all. (Sarah Perez / *TechCrunch*)

Coronavirus may have saved **Facebook** from its fate as a chatroom for old people, this piece argues. There are early signs that young people are returning to the service. (Jael Goldfine / *Paper*)

Facebook’s Menlo Park headquarters have shaped the city. So too would an exodus of employees now that the company is shifting to remote work. (Sarah Emerson / *OneZero*)

Things to do

Stuff to occupy you online during the quarantine.

Listen to Boom / Bust: The Rise and Fall of HQ Trivia. It’s a fun new podcast from *The Ringer* about the company’s dramatic history; I appear on episode two.

Watch all of *Fraggle Rock* on Apple TV+. One of my favorite childhood shows finally has a streaming home.

Check out the launch lineup for HBO Max, which premieres Wednesday. If you already subscribe to HBO Now, as I do, you’re about go get a lot more movies and TV shows for the price.

Subscribe to Alex Kantrowitz's new newsletter about big tech. One of my favorite reporters, Alex announced today he's leaving *BuzzFeed* to go independent. You can sign up to get his new project via email [here](#).

And finally...

I thought my stimulus check came in the mail today and it was a warrent for my arrest

– P ♡ (@queenpalba) May 15, 2020

Talk to us

Send us tips, comments, questions, and YouTube comments critical of the Chinese Communist party: casey@theverge.com and zoe@theverge.com.

<https://www.theverge.com/interface/2020/5/27/21270556/trump-twitter-label-misleading-tweets>