



Michael Shellenberger @shellenberger

Dec 10, 2022 · 40 tweets · [shellenberger/status/1601720455005511680](https://twitter.com/shellenberger/status/1601720455005511680)

1. TWITTER FILES, PART 4

The Removal of Donald Trump: January 7

As the pressure builds, Twitter executives build the case for a permanent ban

On Jan 7, senior Twitter execs:

- create justifications to ban Trump
- seek a change of policy for Trump alone, distinct from other political leaders
- express no concern for the free speech or democracy implications of a ban

This [#TwitterFiles](#) is reported with [@lwoodhouse](#)

For those catching up, please see:

Part 1, where [@mtaibbi](#) documents how senior Twitter executives violated their own policies to prevent the spread of accurate information about Hunter Biden's laptop;



Part 2, where [@bariweiss](#) shows how senior Twitter execs created secret blacklists to “de-amplify” disfavored Twitter users, not just specific tweets;



A screenshot of a tweet from Bari Weiss (@bariweiss) posted on December 9, 2022, at 12:15 AM. The tweet is part of a thread and contains the text: "THREAD: THE TWITTER FILES PART TWO. TWITTER'S SECRET BLACKLISTS." It has received 241.6K likes and 13.3K replies. The interface shows options to reply, copy the link, and read the full conversation on Twitter.

Bari Weiss
@bariweiss · Follow

THREAD: THE TWITTER FILES PART TWO.
TWITTER'S SECRET BLACKLISTS.

12:15 AM · Dec 9, 2022

[Read the full conversation on Twitter](#)

241.6K Likes Reply Copy link

[Read 13.3K replies](#)

And Part 3, where [@mtaibbi](#) documents how senior Twitter execs censored tweets by Trump in the run-up to the Nov 2020 election while regularly engaging with representatives of U.S. government law enforcement agencies.



A screenshot of a tweet from Matt Taibbi (@mtaibbi) posted on December 9, 2022, at 11:04 PM. The tweet is the first in a thread and contains the text: "1. THREAD: The Twitter Files THE REMOVAL OF DONALD TRUMP Part One: October 2020-January 6th". It has received 169.4K likes and 9.4K replies. The interface shows options to reply, copy the link, and read the full conversation on Twitter.

Matt Taibbi ✓
@mtaibbi · Follow

1. THREAD: The Twitter Files
THE REMOVAL OF DONALD TRUMP
Part One: October 2020-January 6th

11:04 PM · Dec 9, 2022

[Read the full conversation on Twitter](#)

169.4K Likes Reply Copy link

[Read 9.4K replies](#)

For years, Twitter had resisted calls to ban Trump.

“Blocking a world leader from Twitter,” it wrote in 2018, “would hide important info... [and] hamper necessary discussion around their words and actions.”



But after the events of Jan 6, the internal and external pressure on Twitter CEO [@jack](#) grows.

Former First Lady [@MichelleObama](#), tech journalist [@karaswisher](#), [@ADL](#), high-tech VC [@ChrisSacca](#), and many others, publicly call on Twitter to permanently ban Trump.

Chris Sacca [@sacca](#) · 6 Jan 2021

You've got blood on your hands, [@jack](#) and Zuck. For four years you've rationalized this terror. Inciting violent treason is not a free speech exercise. If you work at those companies, it's on you too. Shut it down.

12:46 PM · 6 Jan 2021

2,234 Retweets 10,056 Likes

242 2.2K 10K

bobbyg.eth [@bobbyg.eth](#) · 6 Jan 2021
Replying to [@sacca](#) [@jack](#)
Thank you for saying this. ❤️

Rick Zullo [@Rick_Zullo](#) · 6 Jan 2021
Agreed

Deana [@DeanaRM](#) · 6 Jan 2021
Replying to [@sacca](#) [@jack](#)
They haven't just rationalized it, they've monetized it.

Jonathan Greenblatt [@JGreenblattADL](#) · Official

Two bare minimum tweets after the Capitol has been stormed by extremists is too little, too late. [@jack](#), it's overdue to suspend [@realDonaldTrump](#) until his account stops promoting disinformation and inciting violence.

Donald J. Trump [@realDonaldTrump](#) · 33m
I am asking for everyone at the U.S. Capitol to remain peaceful. No violence! Remember, WE are the Party of Law & Order – respect the Law and our great men and women in Blue. Thank you!

Donald J. Trump [@realDonaldTrump](#) · 1h
Please support our Capitol Police and Law Enforcement. They are truly on the side of our Country. Stay peaceful!

Donald J. Trump [@realDonaldTrump](#) · 1h
Mike Pence didn't have the courage to do what should have been done to protect our Country and our Constitution, giving States a chance to certify a corrected set of facts, not the fraudulent or inaccurate ones which they were asked to previously certify. USA demands the truth!

This claim about election fraud is disputed

ADL [@ADL](#) · Jan 6, 2021 · Official
ADL Statement on Violence at the U.S. Capitol Building: [adl.org/news/press-rel.](#)

ADL Statement on Violence at U.S. Capitol Building
New York, NY, January 6, 2021... ADL (Anti-Defamation League) CEO Jonathan Greenblatt issued the following statement regarding today's violence at the U.S. Capitol Building in Washington, D.C.
As ADL has said again and again, extremists must be taken at their word. First there was volatile rhetoric online, then explicit calls to violence and now people are acting on those calls in the nation's capital and flagrantly breaking the law. It must end now.
The President has promoted sedition and incited violence. People assaulting law enforcement officers or breaching government buildings must be arrested and held accountable.
More than anything, what is happening right now at the Capitol is a direct result of the fear and disinformation that has been spread consistently from the Oval Office. President Trump has a responsibility to call for an end to this violence and unrest that he has sowed. His campaign of disinformation is a clear and present danger to our democracy.
But until such time as that happens, social media companies should suspend his accounts ASAP as they would do for anyone else advocating disinformation and promoting violence. It's time.

177 Retweets 26 Quote Tweets 469 Likes

Intelligenceer

It's Time for Social-Media Platforms to Permanently Ban Trump

By Kara Swisher

We have seen the Kraken, and it is Trump. Photo: Warner Bros. Pictures

Donald Trump finally unleashed the Kraken yesterday, although it had little to do with electoral fraud. Instead, it had everything to do with creating frightening chaos and swallowing up enemies whole, which is, for those who

Jonathan Greenblatt [@JGreenblattADL](#) · Official

Two bare minimum tweets after the Capitol has been stormed by extremists is too little, too late. [@jack](#), it's overdue to suspend [@realDonaldTrump](#) until his account stops promoting disinformation and inciting violence.

Donald J. Trump [@realDonaldTrump](#) · 33m
I am asking for everyone at the U.S. Capitol to remain peaceful. No violence! Remember, WE are the Party of Law & Order – respect the Law and our great men and women in Blue. Thank you!

Donald J. Trump [@realDonaldTrump](#) · 1h
Please support our Capitol Police and Law Enforcement. They are truly on the side of our Country. Stay peaceful!

Donald J. Trump [@realDonaldTrump](#) · 1h
Mike Pence didn't have the courage to do what should have been done to protect our Country and our Constitution, giving States a chance to certify a corrected set of facts, not the fraudulent or inaccurate ones which they were asked to previously certify. USA demands the truth!

This claim about election fraud is disputed

ADL [@ADL](#) · Jan 6, 2021 · Official
ADL Statement on Violence at the U.S. Capitol Building: [adl.org/news/press-rel.](#)

ADL Statement on Violence at U.S. Capitol Building
New York, NY, January 6, 2021... ADL (Anti-Defamation League) CEO Jonathan Greenblatt issued the following statement regarding today's violence at the U.S. Capitol Building in Washington, D.C.
As ADL has said again and again, extremists must be taken at their word. First there was volatile rhetoric online, then explicit calls to violence and now people are acting on those calls in the nation's capital and flagrantly breaking the law. It must end now.
The President has promoted sedition and incited violence. People assaulting law enforcement officers or breaching government buildings must be arrested and held accountable.
More than anything, what is happening right now at the Capitol is a direct result of the fear and disinformation that has been spread consistently from the Oval Office. President Trump has a responsibility to call for an end to this violence and unrest that he has sowed. His campaign of disinformation is a clear and present danger to our democracy.
But until such time as that happens, social media companies should suspend his accounts ASAP as they would do for anyone else advocating disinformation and promoting violence. It's time.

177 Retweets 26 Quote Tweets 469 Likes

Dorsey was on vacation in French Polynesia the week of January 4-8, 2021. He phoned into meetings but also delegated much of the handling of the situation to senior execs [@yoyoel](#), Twitter's Global Head of Trust and Safety, and [@vijaya](#) Head of Legal, Policy, & Trust.

As context, it's important to understand that Twitter's staff & senior execs were overwhelmingly progressive.

In 2018, 2020, and 2022, 96%, 98%, & 99% of Twitter staff's political donations went to Democrats.

Replying to @mtaibbi

11. This system wasn't balanced. It was based on contacts. Because Twitter was and is overwhelmingly staffed by people of one political orientation, there were more channels, more ways to complain, open to the left (well, Democrats) than the right. opensecrets.org/orgs/twitter/s

Total	Democrats	% to Dems	Republicans	% to Repubs
\$185,267	\$165,969	99.73%	\$451	0.27%
\$968,749	\$909,431	98.47%	\$14,137	1.53%
\$309,394	\$295,722	96.38%	\$11,100	3.62%

12:02 AM · Dec 3, 2022

112.6K Reply Copy link

Read 3.2K replies

In 2017, Roth tweeted that there were “ACTUAL NAZIS IN THE WHITE HOUSE.”

In April 2022, Roth told a colleague that his goal “is to drive change in the world,” which is why he decided not to become an academic.

Replying to @yoyoel

Yes, that person in the pink hat is clearly a bigger threat to your brand of feminism than ACTUAL NAZIS IN THE WHITE HOUSE.

3:33 PM · Jan 22, 2017

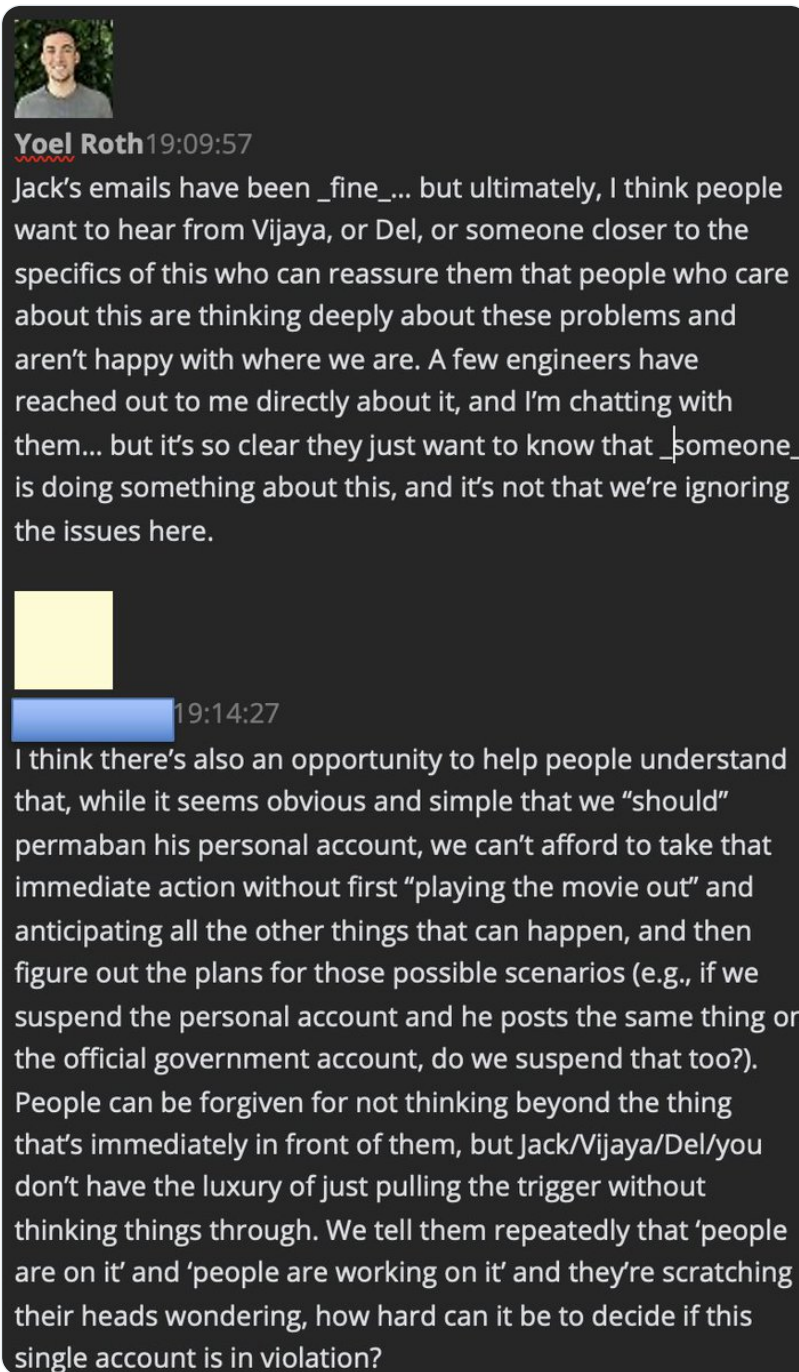
136 Retweets 366 Quote Tweets 217 Likes

Yoel Roth at 2022-04-28 16:05:00

Yeah.... academia is by far the most abusive working environment I've ever been in. The entire system is exploitative in a ton of ways. And also not necessarily productive if your goal is to drive change in the world (which was the main reason I left).

On January 7, @jack emails employees saying Twitter needs to remain consistent in its policies, including the right of users to return to Twitter after a temporary suspension

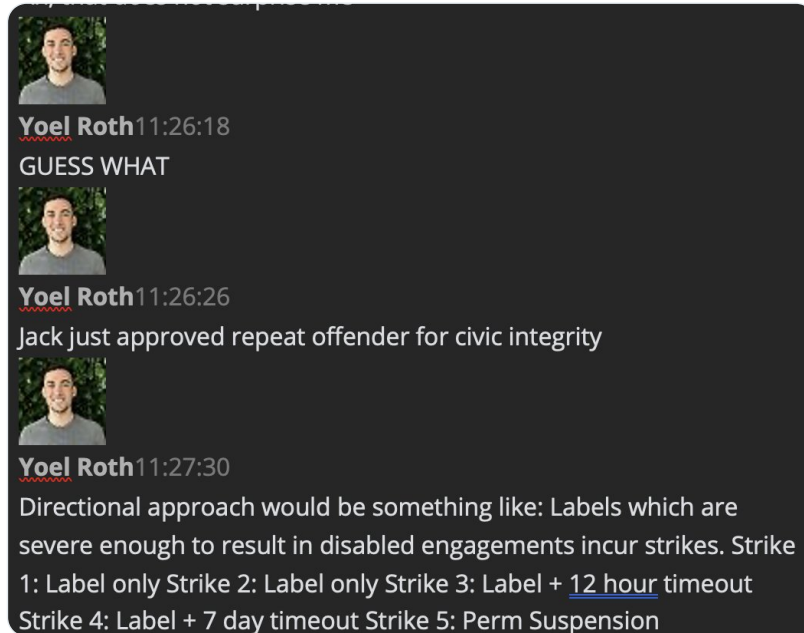
After, Roth reassures an employee that "people who care about this... aren't happy with where we are"



Around 11:30 am PT, Roth DMs his colleagues with news that he is excited to share.

“GUESS WHAT,” he writes. “Jack just approved repeat offender for civic integrity.”

The new approach would create a system where five violations (“strikes”) would result in permanent suspension.

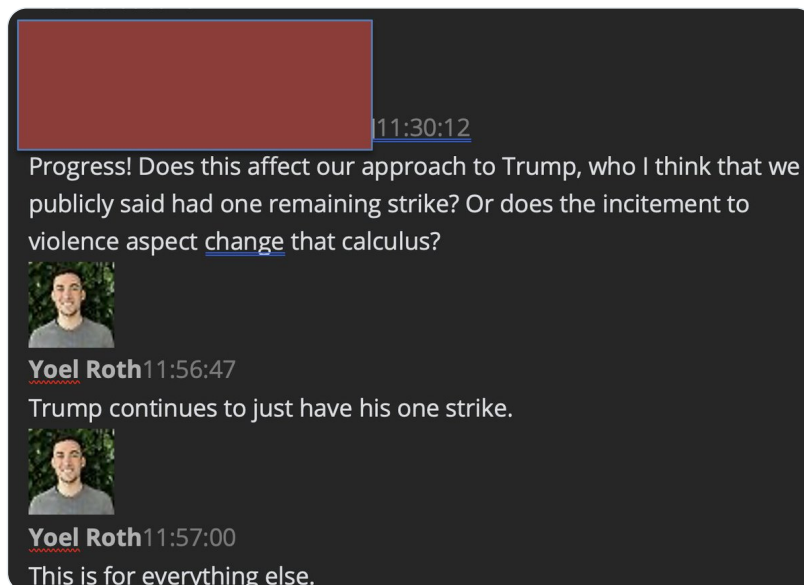


“Progress!” exclaims a member of Roth’s Trust and Safety Team.

The exchange between Roth and his colleagues makes clear that they had been pushing [@jack](#) for greater restrictions on the speech Twitter allows around elections.

The colleague wants to know if the decision means Trump can finally be banned. The person asks, "does the incitement to violence aspect change that calculus?"

Roth says it doesn't. "Trump continues to just have his one strike" (remaining).



Roth's colleague's query about "incitement to violence" heavily foreshadows what will happen the following day.

On January 8, Twitter announces a permanent ban on Trump due to the "risk of further incitement of violence."

Permanent suspension of @realDonaldTrump

By [Twitter Inc.](#)
Friday, 8 January 2021 [Twitter](#) [Facebook](#) [LinkedIn](#) [Share](#)

After close review of recent Tweets from the @realDonaldTrump account and the context around them — specifically how they are being received and interpreted on and off Twitter — we have permanently suspended the account due to the risk of further incitement of violence.

On J8, Twitter says its ban is based on "specifically how [Trump's tweets] are being received & interpreted."

But in 2019, Twitter said it did "not attempt to determine all potential interpretations of the content or its intent."



World Leaders on Twitter: principles & approach

An update on Tweets from world leaders

https://blog.twitter.com/en_us/topics/company/2019/worldleaders2019

After close review of recent Tweets from the @realDonaldTrump account and the context around them — specifically how they are being received and interpreted on and off Twitter — we have permanently suspended the account due to the risk of further incitement of violence.

We focus on the language of reported Tweets and do not attempt to determine all potential interpretations of the content or its intent.

The *only* serious concern we found expressed within Twitter over the implications for free speech and democracy of banning Trump came from a junior person in the organization. It was tucked away in a lower-level Slack channel known as "site-integrity-auto."

08:11:57

This might be an unpopular opinion but one off ad hoc decisions like this that don't appear rooted in policy are imho a slippery slope and reflect an alternatively equally dictatorial problem. This now appears to be a fiat by an online platform CEO with a global presence that can gatekeep speech for the entire world - which seems unsustainable.

"This might be an unpopular opinion but one off ad hoc decisions like this that don't appear rooted in policy are imho a slippery slope... This now appears to be a fiat by an online platform CEO with a global presence that can gatekeep speech for the entire world..."

08:11:57
This might be an unpopular opinion but one off ad hoc decisions like this that don't appear rooted in policy are imho a slippery slope and reflect an alternatively equally dictatorial problem. This now appears to be a fiat by an online platform CEO with a global presence that can gatekeep speech for the entire world - which seems unsustainable.

Twitter employees use the term "one off" frequently in their Slack discussions. Its frequent use reveals significant employee discretion over when and whether to apply warning labels on tweets and "strikes" on users. Here are typical examples.

could have Bounced with a strike. I don't see any scenario where we would decide not the bounce here. Just want to check if there are any concerns, otherwise, I can bounce under RTP and close this one before I go. *FYI* *Account for review: @USAsession* - We are trying to understand the one-off decision here. A few

Friday, November 6th 2020 08.38.11 by [redacted]
there's always AbuseOne-Off

Friday, November 6th 2020 08.38.33 by [redacted]
Bounce. (One-off) is the option

Recall from [#TwitterFiles2](#) by [@bariweiss](#) that, according to Twitter staff, "We control visibility quite a bit. And we control the amplification of your content quite a bit. And normal people do not know how much we do."



Bari Weiss
@bariweiss · Follow

Replying to @bariweiss

11. "We control visibility quite a bit. And we control the amplification of your content quite a bit. And normal people do not know how much we do," one Twitter engineer told us. Two additional Twitter employees confirmed.


12:48 AM · Dec 9, 2022


70.9K Reply Copy link


Read 804 replies


Twitter employees recognize the difference between their own politics & Twitter's Terms of Service (TOS), but they also engage in complex interpretations of content in order to stamp out prohibited tweets, as a series of exchanges over the "[#stopthesteal](#)" hashtag reveal.


Thursday, January 7th 2021


 10:57:37
hey - hope you're doing ok and were able to get some sleep. can we, or have we already discussed, blocking the #stopthesteal hashtag? it's furthering fake news in a dangerous way from what i can tell.



Yoel Roth 11:55:55
Hey there!



Yoel Roth 11:56:20
We're proactively surfacing that content for review under Civic Integrity — and it should be blocked from typeahead and trends. If you're seeing it in either of the latter 2, let me know ASAP and PT can fix.

 11:57:53
i just saw the hashtag and then did a search on it and all the content was appalling. so i don't think it's showing up elsewhere. thank you for all you're doing


Yoel Roth 12:09:46
Hmm

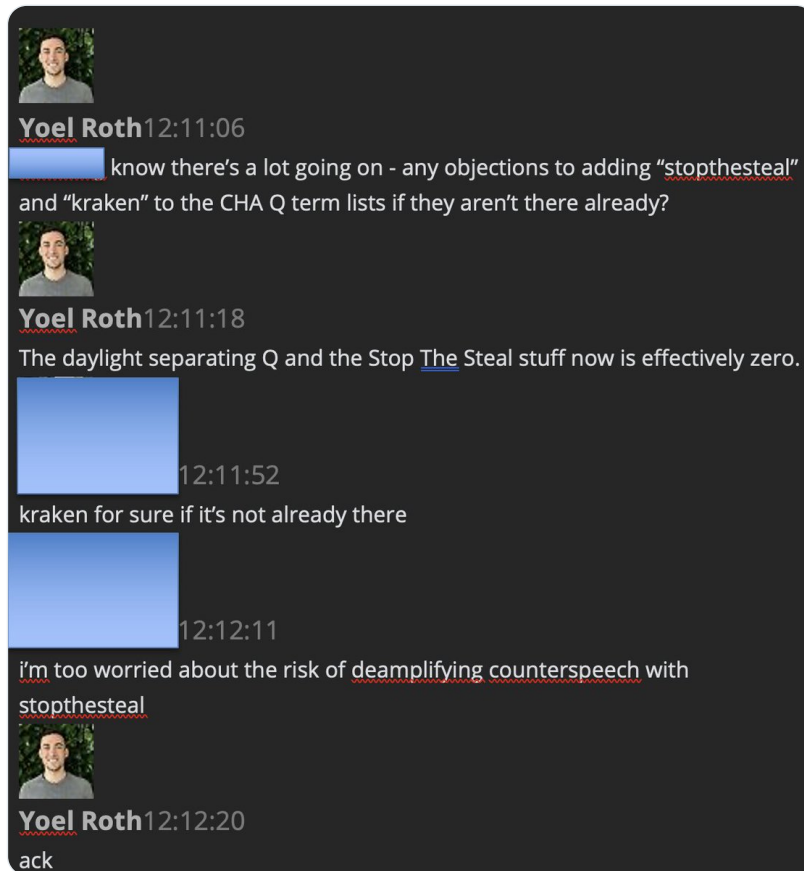

Yoel Roth 12:09:49
I'll flag to the election squad

 12:10:28
remember that what i find appalling and what is a TOS vio are not an exact overlap


Yoel Roth 12:10:39
heh

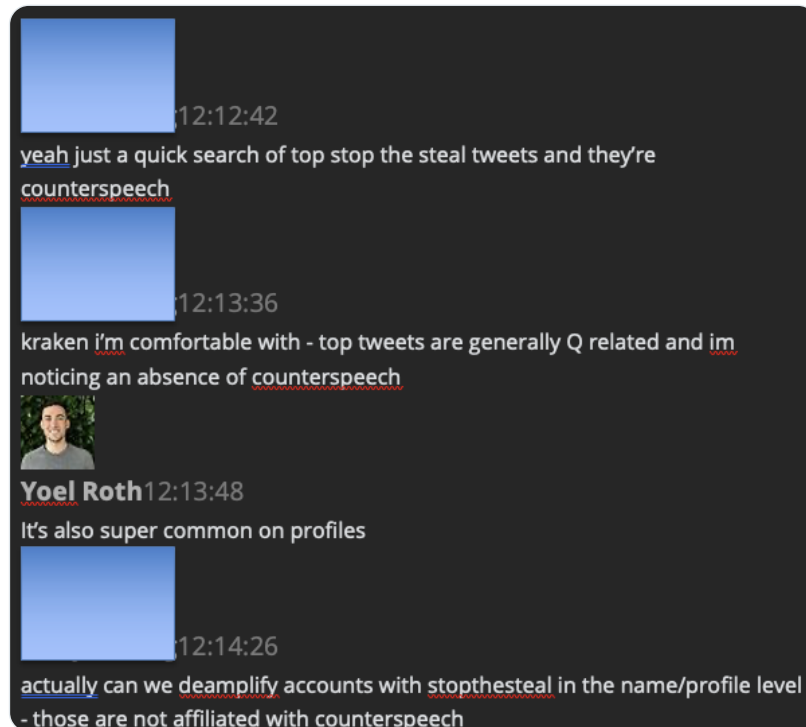
Roth immediately DMs a colleague to ask that they add "stopthesteal" & [QAnon conspiracy term] "kraken" to a blacklist of terms to be deamplified.

Roth's colleague objects that blacklisting "stopthesteal" risks "deamplifying counterspeech" that validates the election.

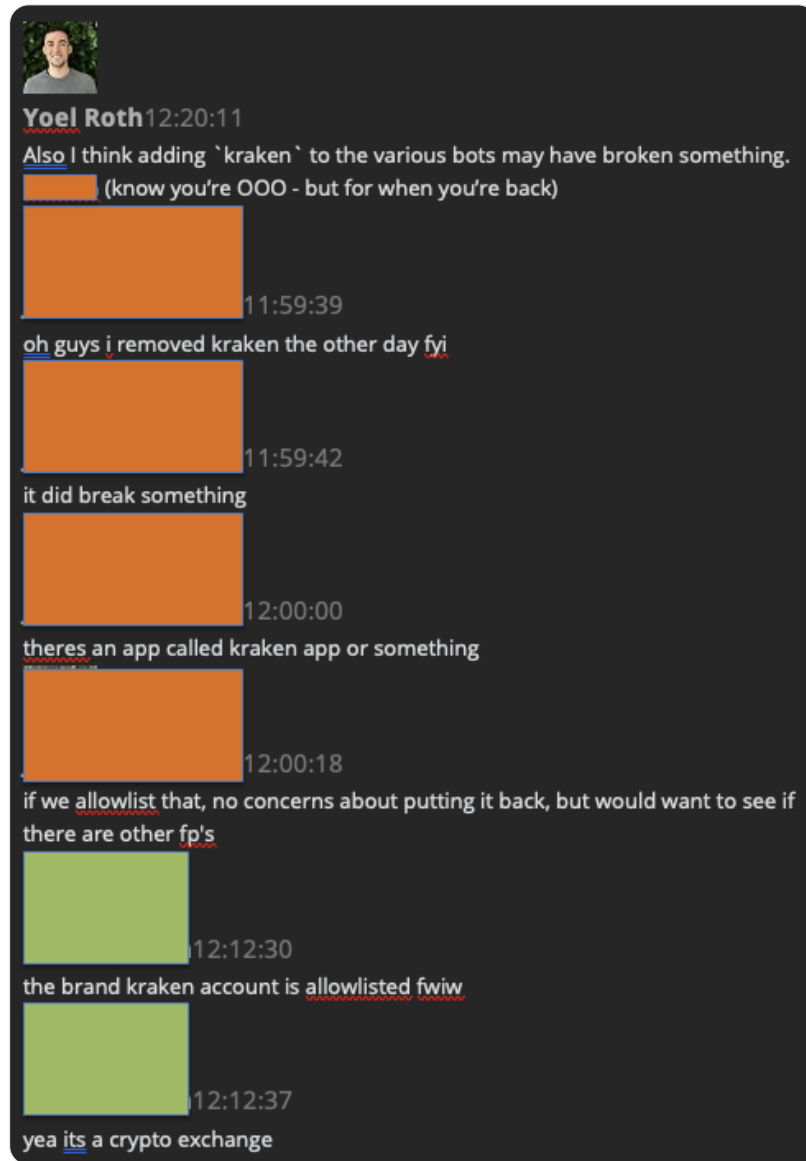


Indeed, notes Roth's colleague, "a quick search of top stop the steal tweets and they're counterspeech"

But they quickly come up with a solution: "deamplify accounts with stopthesteal in the name/profile" since "those are not affiliated with counterspeech"



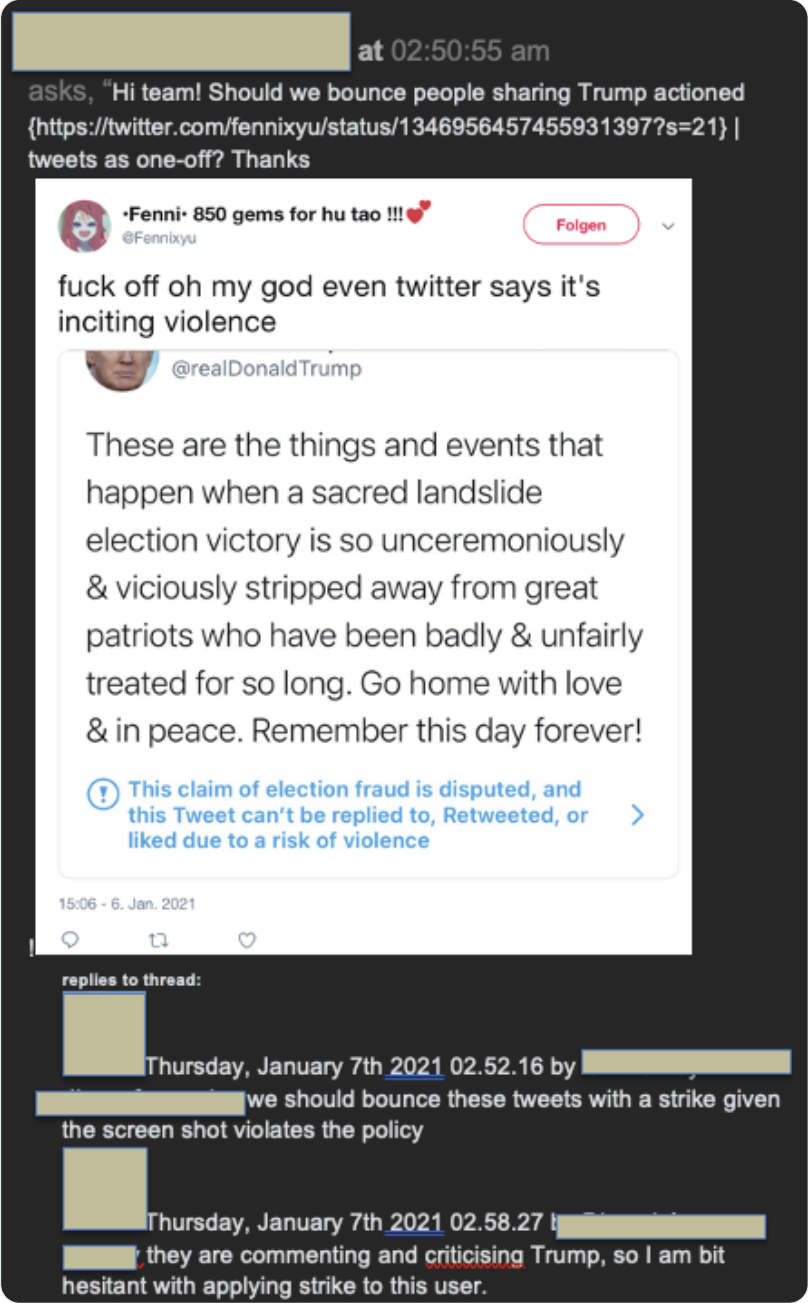
But it turns out that even blacklisting "kraken" is less straightforward than they thought. That's because kraken, in addition to being a QAnon conspiracy theory based on the mythical Norwegian sea monster, is also the name of a cryptocurrency exchange, and was thus "allowlisted"



Employees struggle with whether to punish users who share screenshots of Trump's deleted J6 tweets

"we should bounce these tweets with a strike given the screen shot violates the policy"

"they are criticising Trump, so I am bit hesitant with applying strike to this user"



What if a user dislikes Trump *and* objects to Twitter's censorship? The tweet still gets deleted. But since the *intention* is not to deny the election result, no punishing strike is applied.

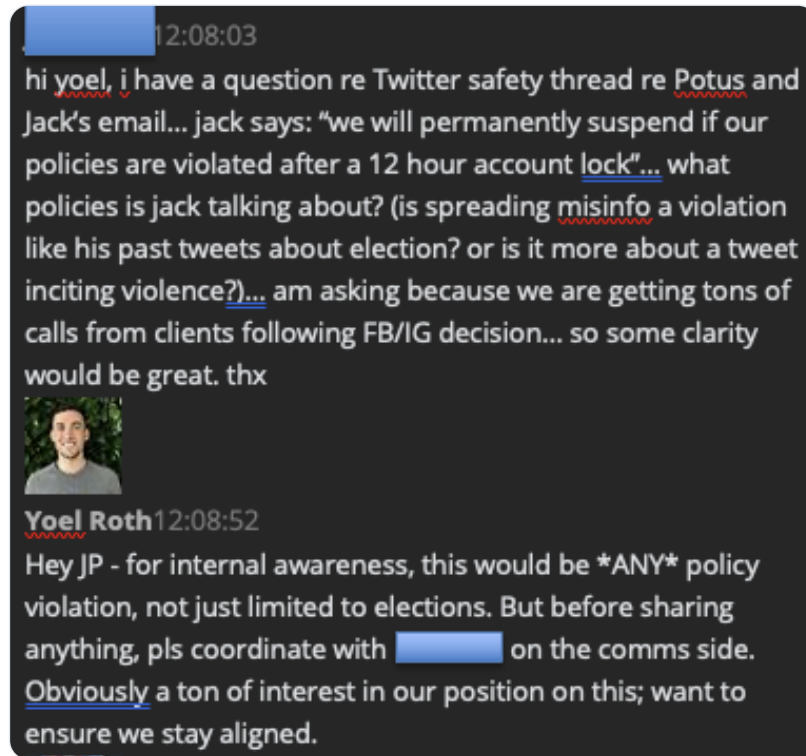
"if there are instances where the intent is unclear please feel free to raise"



Around noon, a confused senior executive in advertising sales sends a DM to Roth.

Sales exec: "jack says: 'we will permanently suspend [Trump] if our policies are violated after a 12 hour account lock'... what policies is jack talking about?"

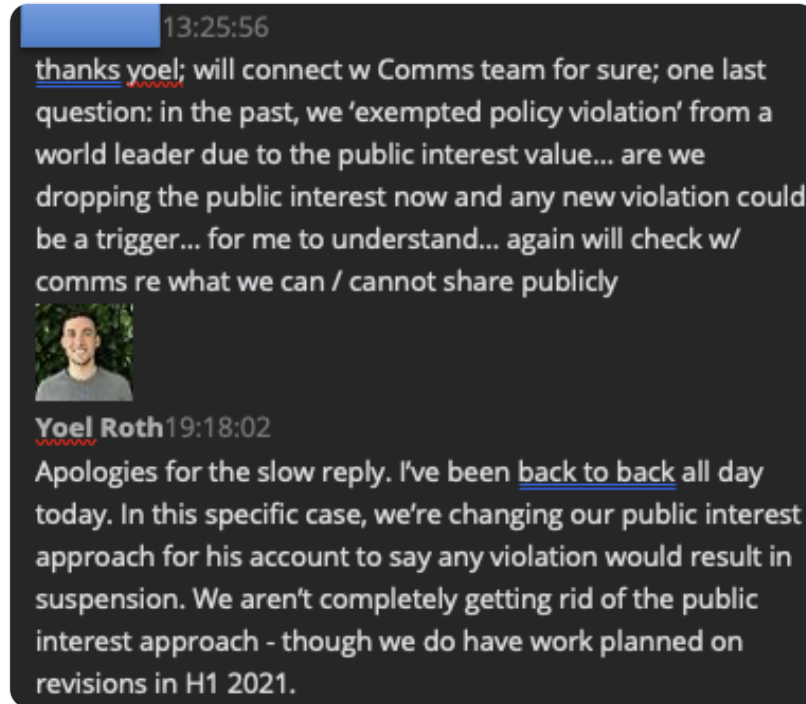
Roth: "*ANY* policy violation"



What happens next is essential to understanding how Twitter justified banning Trump.

Sales exec: "are we dropping the public interest [policy] now..."

Roth, six hours later: "In this specific case, we're changing our public interest approach for his account..."



The ad exec is referring to Twitter's policy of "Public-interest exceptions," which allows the content of elected officials, even if it violates Twitter rules, "if it directly contributes to understanding or discussion of a matter of public concern"



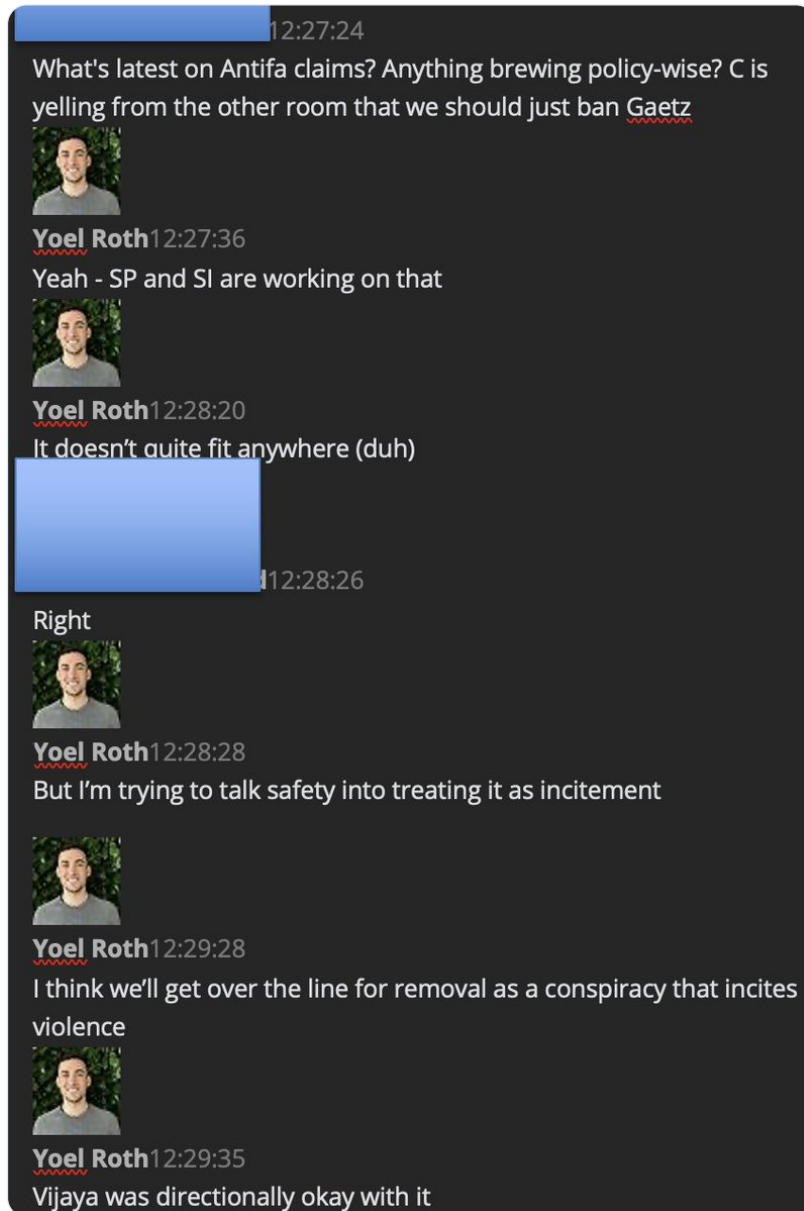
Public-interest exceptions to enforcement of Twitter rules
Learn why we make certain exceptions, under what circumstances, and how we balance risk of harm vs. the public interest.
<https://help.twitter.com/en/rules-and-policies/public-interest>

At present, we limit exceptions to one critical type of public-interest content—Tweets from elected and government officials—given the significant public interest in knowing and being able to discuss their actions and statements.

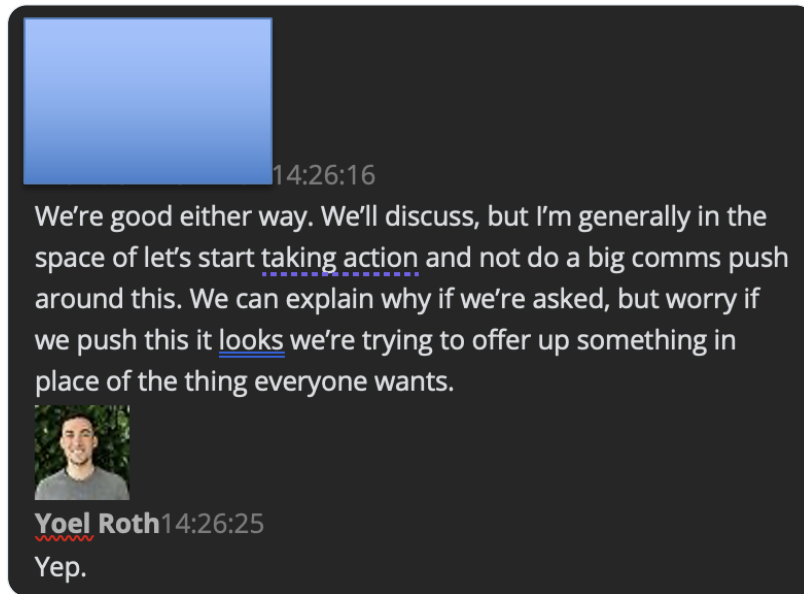
Roth pushes for a permanent suspension of Rep. Matt Gaetz even though it “doesn’t quite fit anywhere (duh)”

It's a kind of test case for the rationale for banning Trump.

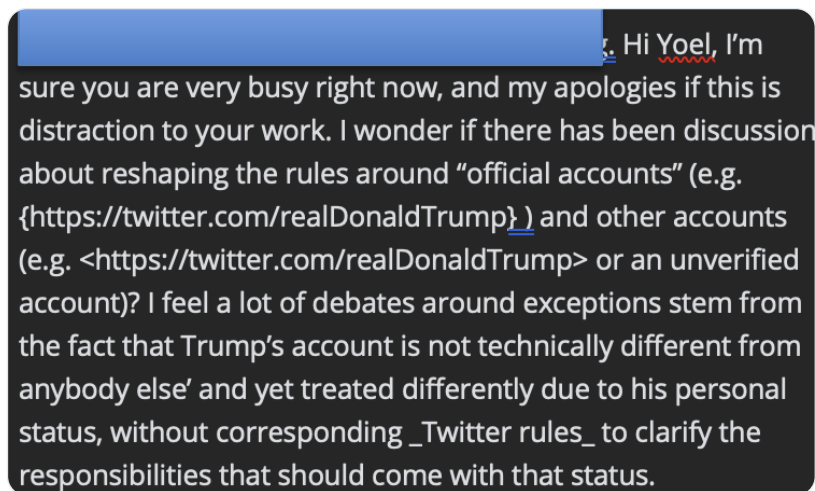
“I’m trying to talk [Twitter’s] safety [team] into... removal as a conspiracy that incites violence.”



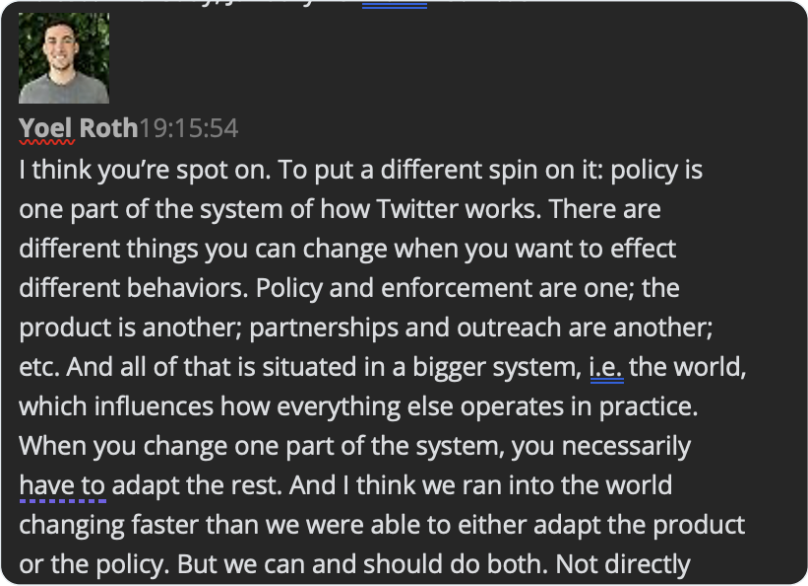
Around 2:30, comms execs DM Roth to say they don't want to make a big deal of the QAnon ban to the media because they fear "if we push this it looks we're trying to offer up something in place of the thing everyone wants," meaning a Trump ban.



That evening, a Twitter engineer DMs to Roth to say, "I feel a lot of debates around exceptions stem from the fact that Trump's account is not technically different from anybody else' and yet treated differently due to his personal status, without corresponding _Twitter rules_."



Roth's response hints at how Twitter would justify deviating from its longstanding policy. "To put a different spin on it: policy is one part of the system of how Twitter works... we ran into the world changing faster than we were able to either adapt the product or the policy."



The evening of January 7, the same junior employee who expressed an "unpopular opinion" about "ad hoc decisions... that don't appear rooted in policy," speaks up one last time before the end of the day.



Earlier that day, the employee wrote, "My concern is specifically surrounding the unarticulated logic of the decision by FB. That space fills with the idea (conspiracy theory?) that all... internet moguls... sit around like kings casually deciding what people can and cannot see."

Thursday, January 7th 2021 08.51.27 by
My concern is specifically surrounding the unarticulated logic of the decision by FB. That's the space that fills with the idea (conspiracy theory?) that all social media heads and internet moguls at every layer sit around like kings casually deciding what people can and cannot see, and it's unhelpful to the internet ecosystem as a whole. Again this is IMHO only.

The employee notes, later in the day, "And Will Oremus noticed the inconsistency too...," linking to an article for OneZero at Medium called, "Facebook Chucked Its Own Rulebook to Ban Trump."



"The underlying problem," writes [@WillOremus](#), is that "the dominant platforms have always been loath to own up to their subjectivity, because it highlights the extraordinary, unfettered power they wield over the global public square..."

"... and places the responsibility for that power on their own shoulders... So they hide behind an ever-changing rulebook, alternately pointing to it when it's convenient and shoving it under the nearest rug when it isn't."



"Facebook's suspension of Trump now puts Twitter in an awkward position. If Trump does indeed return to Twitter, the pressure on Twitter will ramp up to find a pretext on which to ban him as well."

Indeed. And as [@bariweiss](#) will show tomorrow, that's exactly what happened.

/END



...