

Pinned Tweet



Michael Shellenberger

@ShellenbergerMD

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;



Matt Taibbi

@mtaibbi - Dec 2

1. Thread: THE TWITTER FILES [Show this thread](#)

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

Quote Tweet



Bari Weiss

@bariweiss · Dec 8

THREAD: THE TWITTER FILES PART TWO. TWITTER’S SECRET BLACKLISTS. [Show this thread](#)

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.



Matt Taibbi

@mtaibbi · Dec 9

1. THREAD: The Twitter Files THE REMOVAL OF DONALD TRUMP Part One: October 2020-January 6th [Show this thread](#)

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.”



Twitter Public Policy

@Policy · Jan 5, 2018

Blocking a world leader from Twitter or removing their controversial Tweets would hide important information people should be able to see and debate. It would also not silence that leader, but it would certainly hamper necessary discussion around their words and actions.

[Show this thread](#)

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

Follow

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** @rsg · 6 Jan 2021
Replying to @sacca @jack
Thank you for saying this. ❤️
2 35
-  **Rick Zullo** @Rick_Zullo · 6 Jan 2021
Agreed
2
-  **Deana** @DeanaRM · 6 Jan 2021
Replying to @sacca @jack
They haven't just rationalized it, they've monetized it.
1 17 397

2 / 2



Seeing the gulf between the responses to yesterday's riot and this summer's peaceful protests and the larger movement for racial justice is so painful. It hurts. And I cannot think about moving on or turning the page until we reckon with the reality of what we saw yesterday. True progress will be possible only once we acknowledge that this disconnect exists and take steps to repair it. And that also means coming to grips with the reality that millions voted for a man so obviously willing to burn our democracy down for his own ego.

I hurt for our country. And I wish I had all the solutions to make things better. I wish I had the confidence that people who know better will act like it for more than a news cycle or two. All I know is that now is a time for true patriotism. Now is the time for those who voted for this president to see the reality of what they've supported—and publicly and forcefully rebuke him and the actions of that mob. [Now is the time for Silicon Valley companies to stop enabling this monstrous behavior—and go even further than they have already by permanently banning this man from their platforms and putting in place policies to prevent their technology from being used by the nation's leaders to fuel insurrection.](#)

And if we have any hope of improving this nation, now is the time for swift and serious consequences for the failure of leadership that led to yesterday's shame.

Thankfully, even in the darkness, there are glimmers of hope. It's something I imagine Reverend Warnock has preached about before—and I'm still heartened beyond belief that he's headed to Washington. I'm glad his fellow Georgian, Jon Ossoff is, too, and that together they'll help give control of Congress back to the only party that's shown that it can put our democracy above its own short-term political fortunes. I pray that every American, especially those who disagree with them, will give our new Congress, President-Elect Biden, and Vice-President-Elect Harris the chance to lead us in a better direction.


But make no mistake: The work of putting America back together, of truly repairing what is broken, isn't the work of any individual politician or political party. It's up to each of us to do our part. To reach out. To listen. And to hold tight to the truth and values that have always led this country forward. It will be an uncomfortable, sometimes painful process. But if we enter into it with an honest and unwavering love of our country, then maybe we can finally start to heal.


 **Michelle Obama** 
@MichelleObama


12:49 PM · Jan 7, 2021

7,262 Retweets 416 Quote Tweets
47.5K Likes

Tweet your reply

 **Stuart D.** @S... · Jan 7, 2021
Replying to @MichelleObama
"Peaceful BLM protests" this past summer?? Seriously? Is that what you call arson, looting and mayhem? And I dont recall any police shoot,let alone kill, any protestor. Not defending yesterday,but please don't underestimate the intelligence of Americans
nypost.com/2021/01/08/fem...

 **Lynda...** @Lyn... · Jan 7, 2021
Replying to @MichelleObama
I'm with my forever FLOTUS. And also, with Porch Guy.

 **Amanda G.** @A... · Jan 7, 2021
Replying to @MichelleObama
Thank you.



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!

159.6K

98.4K

236.4K



Donald J. Trump ✓ @realDonaldTrump · 1h

Please support our Capitol Police and Law Enforcement. They are truly on the side of our Country. Stay peaceful!

171.6K

95.3K

272.7K



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 spewed 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.



12:50 PM · Jan 6, 2021

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.



Matt Taibbi

@mtaibbi · Dec 2

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.

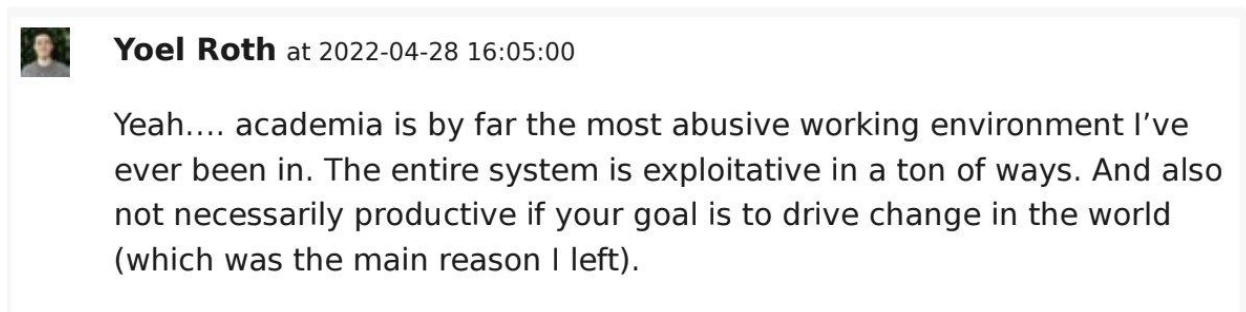
...[opensecrets.org/orgs/twitter/s...](https://www.opensecrets.org/orgs/twitter/s...) [Show this thread](#)

Contributions by Party of Recipient

Cycle	Total	Democrats	% to Dems	Republicans	% to Repubs
2022	\$185,267	\$165,969	99.73%	\$451	0.27%
2020	\$968,749	\$909,431	98.47%	\$14,137	1.53%
2018	\$309,394	\$295,722	96.38%	\$11,100	3.62%

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.



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"



Yoel Roth 19:09:57

Jack's emails have been _fine_... but ultimately, I think people want to hear from Vijaya, or Del, or someone closer to the specifics of this who can reassure them that people who care about this are thinking deeply about these problems and aren't happy with where we are. A few engineers have reached out to me directly about it, and I'm chatting with them... but it's so clear they just want to know that someone is doing something about this, and it's not that we're ignoring the issues here.



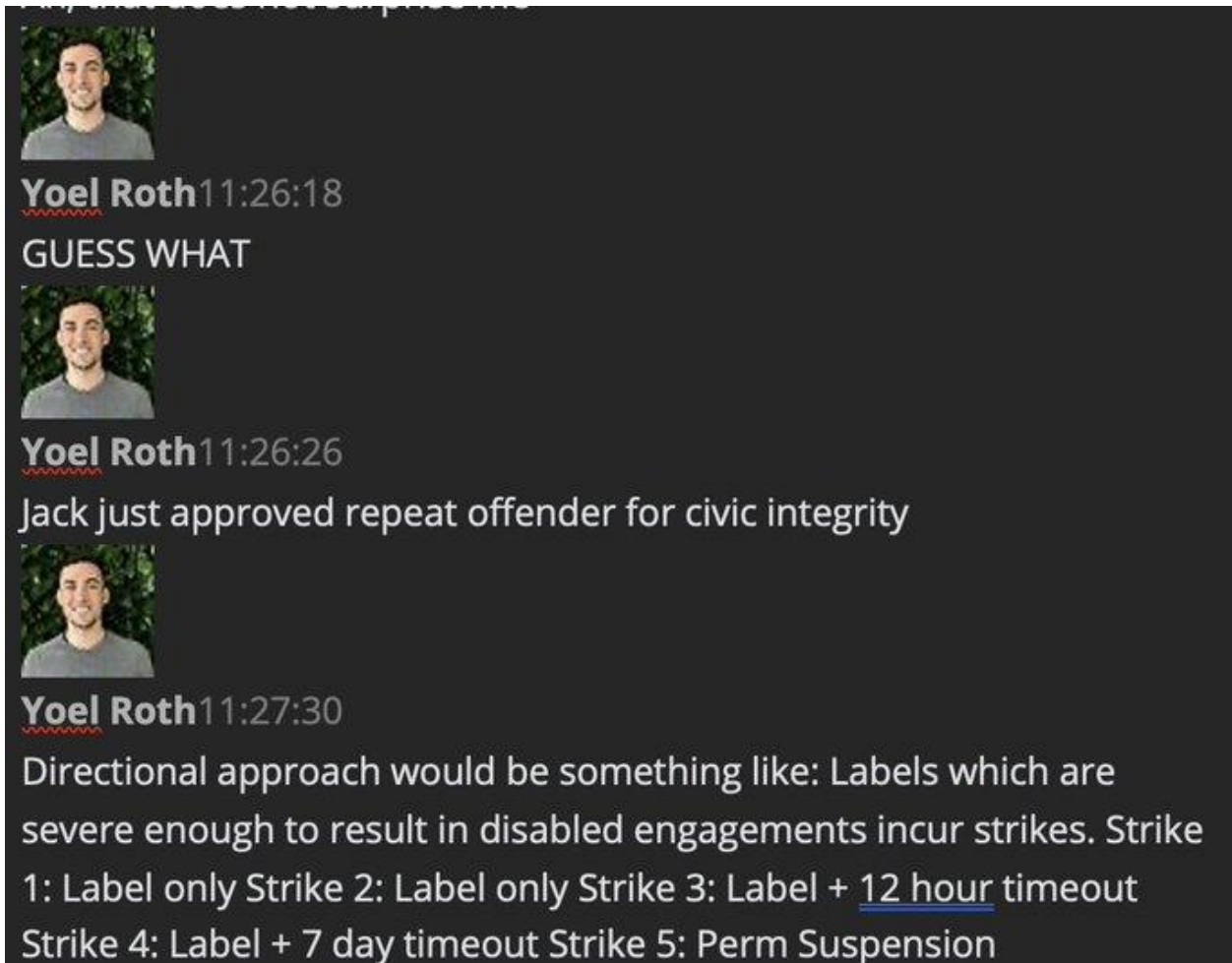
19:14:27

I think there's also an opportunity to help people understand that, while it seems obvious and simple that we "should" permaban his personal account, we can't afford to take that immediate action without first "playing the movie out" and anticipating all the other things that can happen, and then figure out the plans for those possible scenarios (e.g., if we suspend the personal account and he posts the same thing on the official government account, do we suspend that too?). People can be forgiven for not thinking beyond the thing that's immediately in front of them, but Jack/Vijaya/Del/you don't have the luxury of just pulling the trigger without thinking things through. We tell them repeatedly that 'people are on it' and 'people are working on it' and they're scratching their heads wondering, how hard can it be to decide if this single account is in violation?

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.

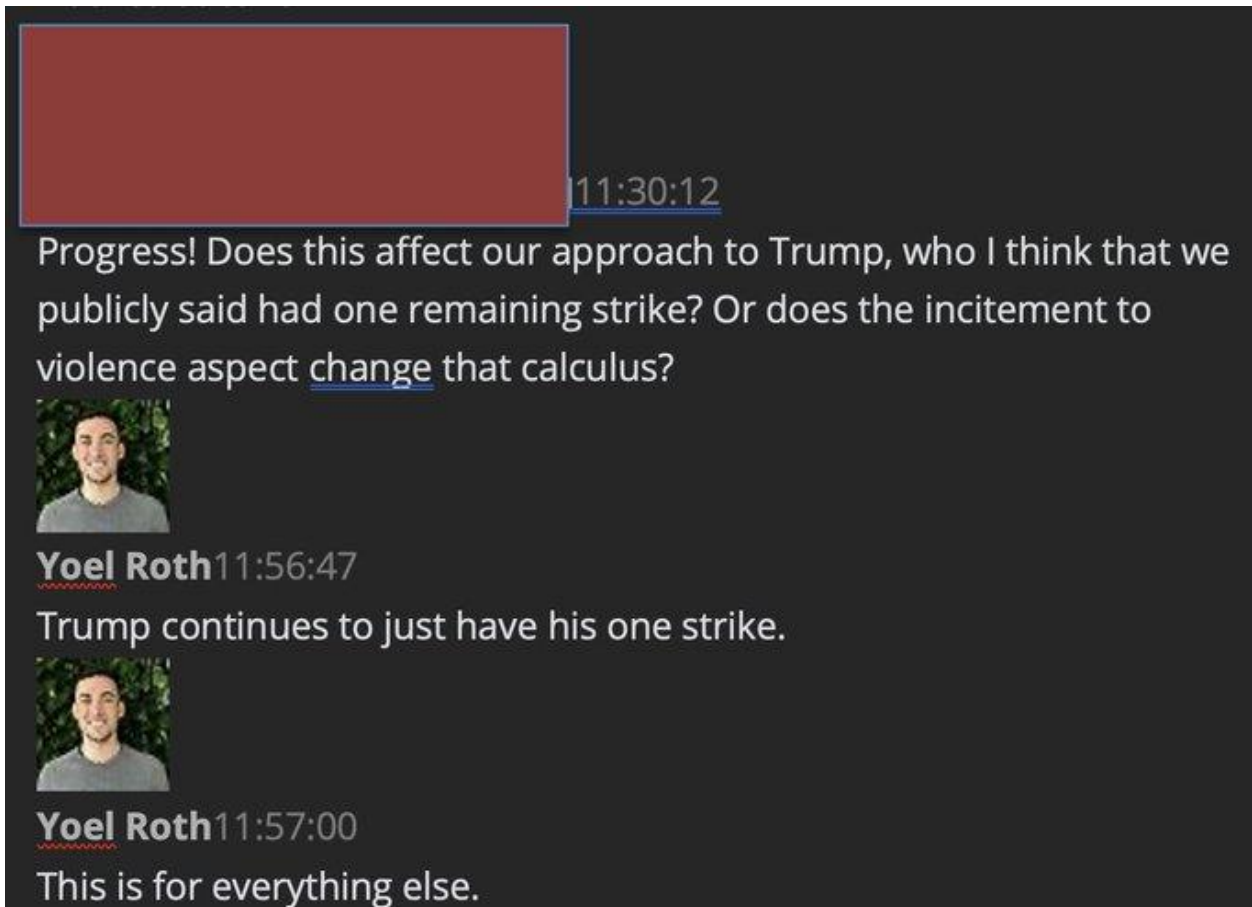


“Progress!” exclaims a member of Roth’s Trust and Safety Team.

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    

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."

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

(See link)

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.

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

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. *FYIs* *Account for review: @USAsession* - We are trying to understand the one-off decision here. A few

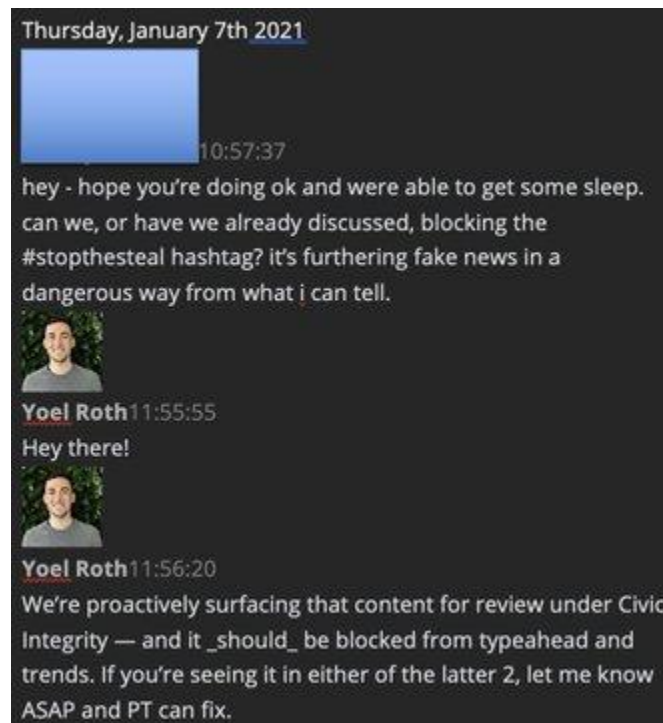
Recall from #TwitterFiles2 by @bariweiss that, according to Twitter staff, "We control visibility quite a bit. And we control the amplification of your content quite a bit. And normal people do not know how much we do."

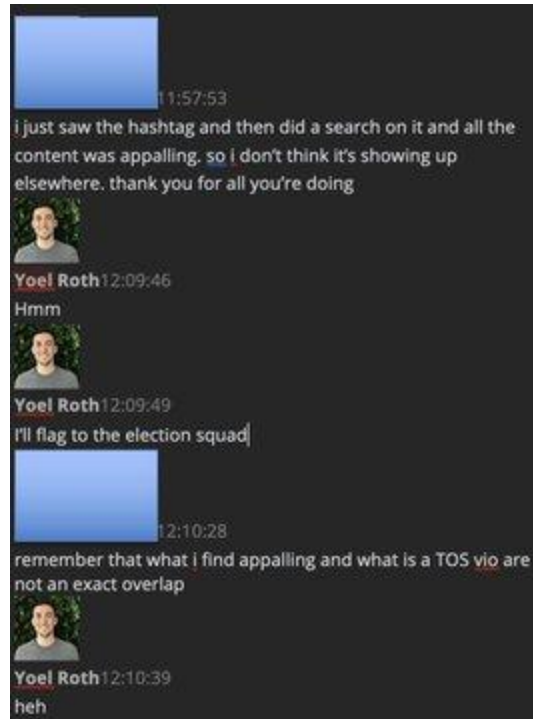
Bari Weiss

@bariweiss · Dec 8

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. [Show this thread](#)

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.





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.



Yoel Roth 12:11:06

know there's a lot going on - any objections to adding "stopthesteal" and "kraken" to the CHA Q term lists if they aren't there already?



Yoel Roth 12:11:18

The daylight separating Q and the Stop The Steal stuff now is effectively zero.



12:11:52

kraken for sure if it's not already there



12:12:11

i'm too worried about the risk of deamplifying counterspeech with stopthesteal



Yoel Roth 12:12:20

ack

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"



12:12:42

yeah just a quick search of top stop the steal tweets and they're
counterspeech



12:13:36

kraken i'm comfortable with - top tweets are generally Q related and im
noticing an absence of counterspeech



Yoel Roth 12:13:48

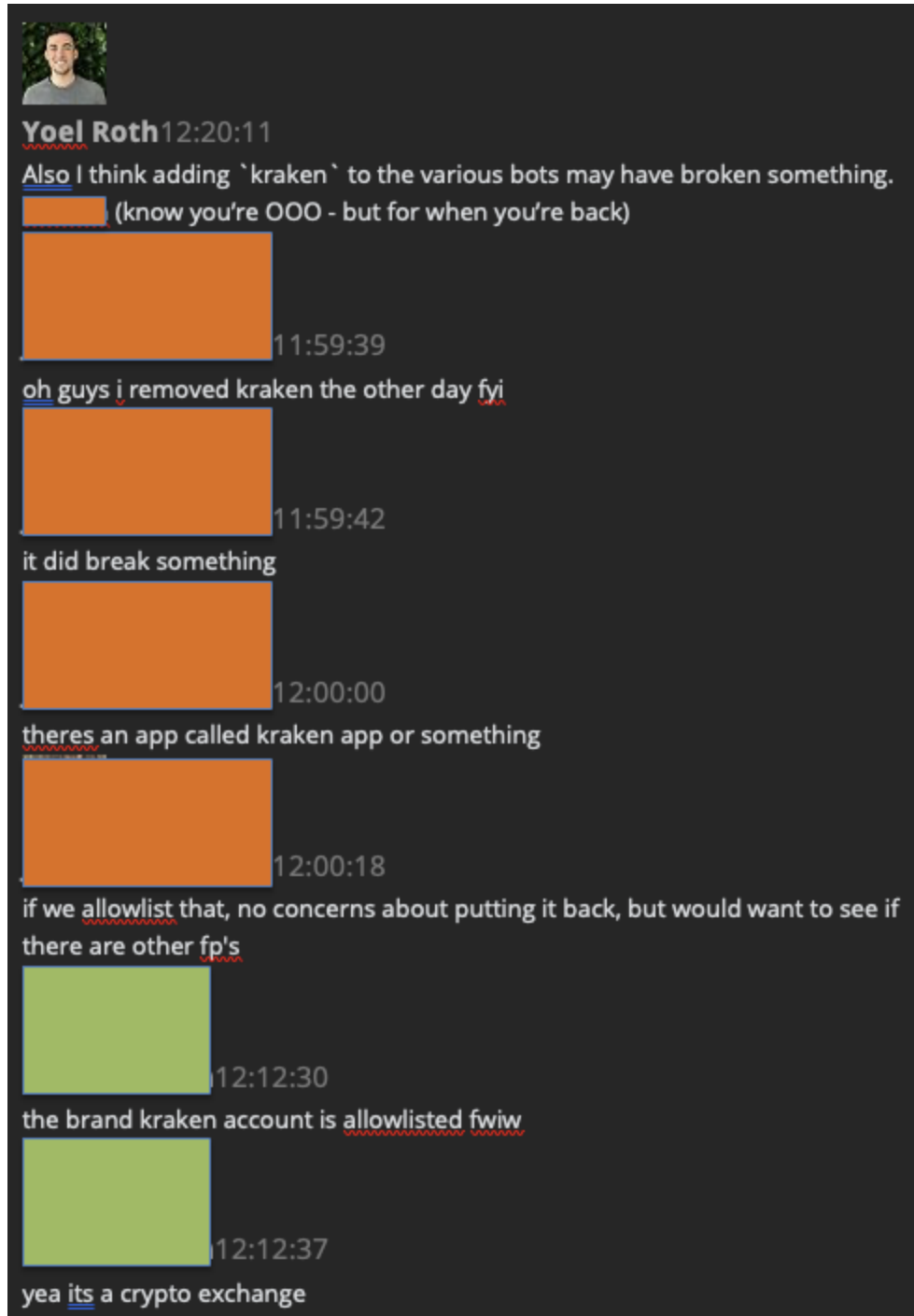
It's also super common on profiles



12:14:26

actually can we deamplify accounts with stopthesteal in the name/profile level
- 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"


[redacted] at 02:50:55 am
asks, "Hi team! Should we bounce people sharing Trump actioned {<https://twitter.com/fennixyu/status/1346956457455931397?s=21>} | tweets as one-off? Thanks

 •Fenni• 850 gems for hu tao !!! ❤️
@Fennixyu Folgen

fuck off oh my god even twitter says it's inciting violence


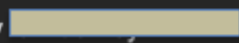
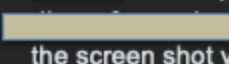
 @realDonaldTrump


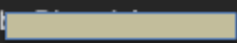

These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away from great patriots who have been badly & unfairly treated for so long. Go home with love & in peace. Remember this day forever!

 This claim of election fraud is disputed, and this Tweet can't be replied to, Retweeted, or liked due to a risk of violence

15:06 - 6. Jan. 2021

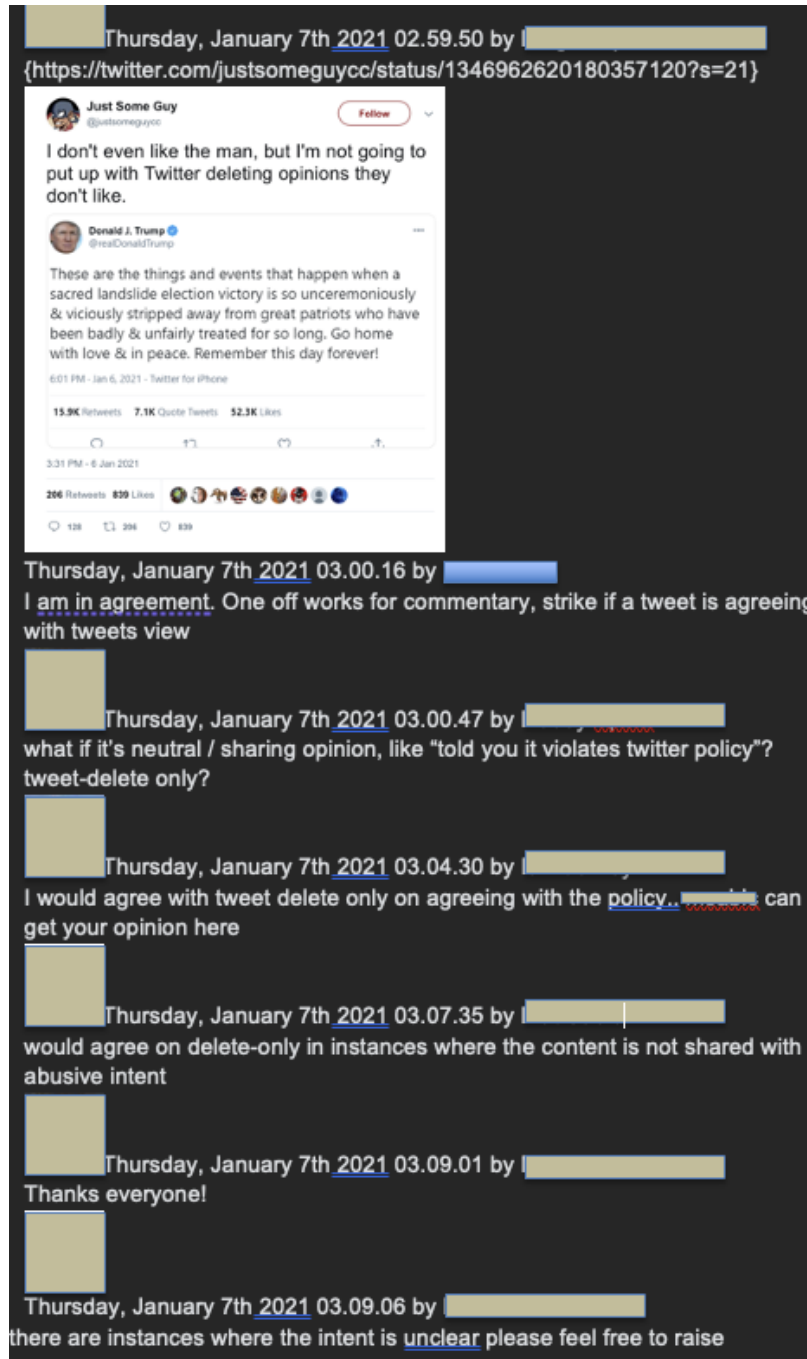
replies to thread:

 Thursday, January 7th 2021 02.52.16 by 
 we should bounce these tweets with a strike given the screen shot violates the policy

 Thursday, January 7th 2021 02.58.27 by 
 they are commenting and 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.

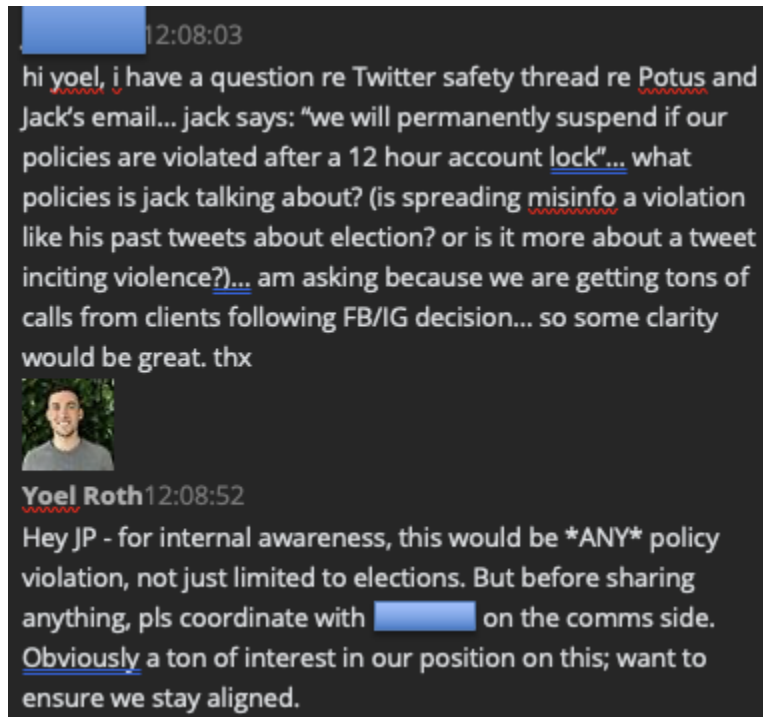
"if there are instances where the intent is unclear please feel free to raise"



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?"

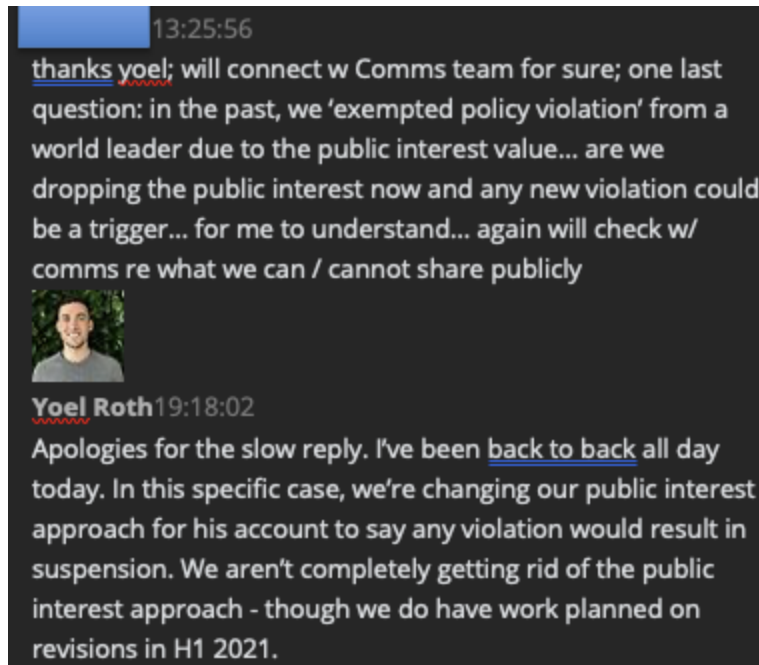
Roth: "*ANY* policy violation"



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"

<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."

12:27:24

What's latest on Antifa claims? Anything brewing policy-wise? C is yelling from the other room that we should just ban Gaetz



Yoel Roth 12:27:36

Yeah - SP and SI are working on that



Yoel Roth 12:28:20

It doesn't quite fit anywhere (duh)



12:28:26

Right



Yoel Roth 12:28:28

But I'm trying to talk safety into treating it as incitement



Yoel Roth 12:29:28

I think we'll get over the line for removal as a conspiracy that incites violence



Yoel Roth 12:29:35

Vijaya was directionally okay with it

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.



14:26:16

We're good either way. We'll discuss, but I'm generally in the space of let's start taking action and not do a big comms push around this. We can explain why if we're asked, but worry if we push this it looks we're trying to offer up something in place of the thing everyone wants.



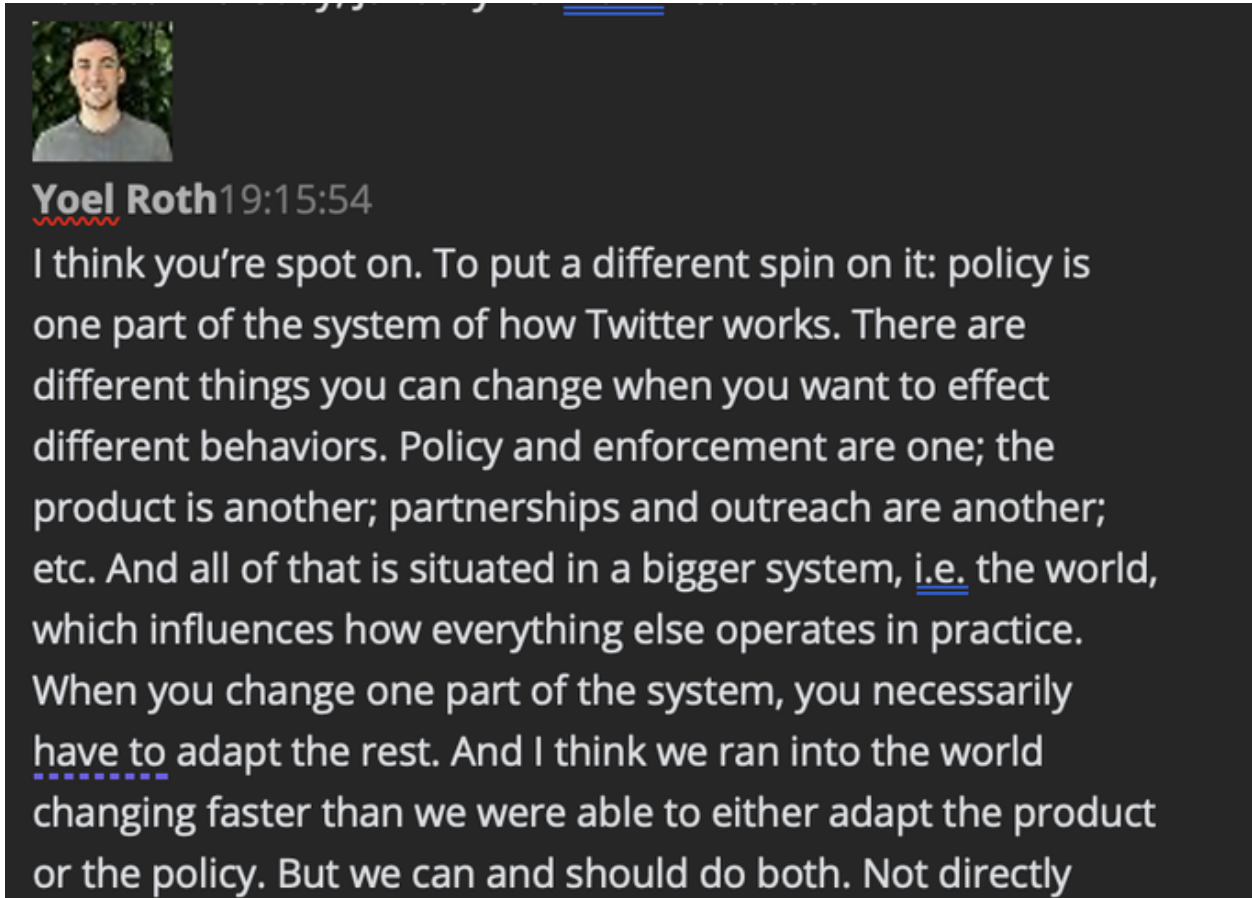
Yoel Roth 14:26:25

Yep.

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_."

Hi Yoel, I'm
sure you are very busy right now, and my apologies if this is
distraction to your work. I wonder if there has been discussion
about reshaping the rules around "official accounts" (e.g.
{<https://twitter.com/realDonaldTrump>}) and other accounts
(e.g. <<https://twitter.com/realDonaldTrump>> or an unverified
account)? 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_ to clarify the
responsibilities that should come with that status.

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.

<https://twitter.com/ShellenbergerMD/status/1601738653805387779>

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." onezero.medium.com

[Facebook Chucked Its Own Rulebook to Ban Trump](#)

[The move is a reminder of social platforms' power over online speech – and the inconsistency with which they wield it](#)

"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