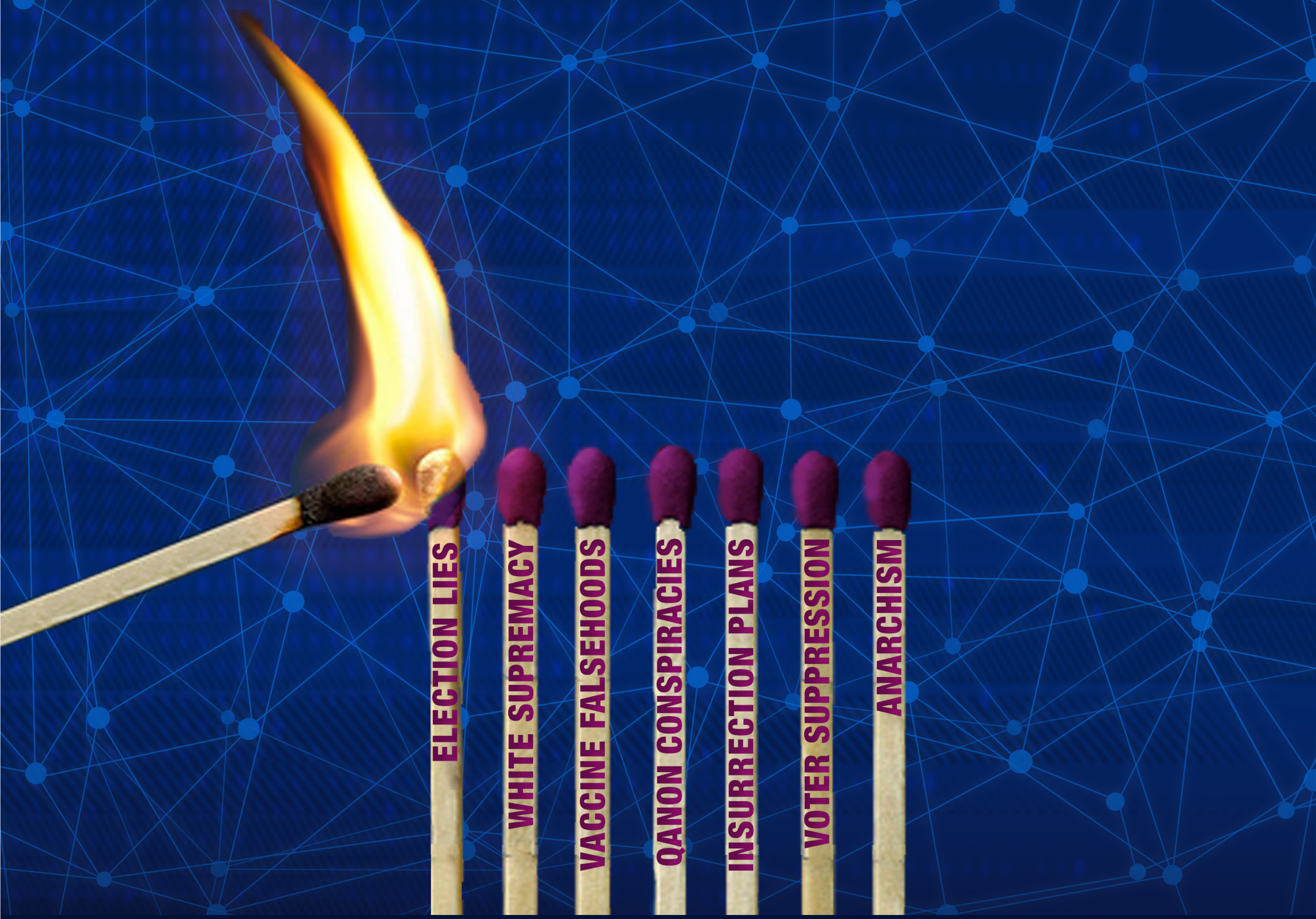
Fueling the Fire:

How Social Media Intensifies U.S. Political Polarization – And What Can Be Done About It

PAUL M. BARRETT, JUSTIN HENDRIX, and J. GRANT SIMS



NYU | STERN

Center for Business
and Human Rights

September 2021

Contents

Executive Summary 1

Introduction 3

Part 1: Assessing the Research on Social Media and Polarization 6

Sidebar: Race, Realignment, and Cable TV 8

Part 2: Asymmetric Polarization and its Consequences 15

Sidebar: Lessons From Abroad 18

Part 3: Conclusion and Recommendations 22

Endnotes 27

Appendix: People Interviewed for This Report 29

Authors

Paul M. Barrett is deputy director of the NYU Stern Center for Business and Human Rights and an adjunct professor at the NYU School of Law.

Justin Hendrix is an associate research scientist and adjunct professor at NYU Tandon School of Engineering and the CEO and editor of Tech Policy Press, a non-profit media venture concerned with the intersection of technology and democracy.

J. Grant Sims is a Ropes & Gray research fellow at the NYU Stern Center for Business and Human Rights.

Acknowledgments

We are grateful to Craig Newmark Philanthropies and the Open Society Foundations for their continued support of our work on the technology sector.

Executive Summary

Some critics of the social media industry contend that widespread use of Facebook, Twitter, and YouTube has contributed to increased political polarization in the United States. But Facebook, the largest social media platform, has disputed this contention, saying that it is unsupported by social science research. Determining whether social media plays a role in worsening partisan animosity is important because political polarization has pernicious consequences.

“Determining the actual relationship between social media and partisan animosity is important and urgent because the current extreme level of divisiveness in the United States is having pernicious consequences.”

In the U.S., where partisan divisiveness has reached new extremes, these consequences include declining trust in fellow citizens and major institutions; erosion of democratic norms like respect for elections; loss of faith in the existence of commonly held facts; and political violence such as the January 6, 2021, insurrection on Capitol Hill.

This report analyzes the evidence bearing on social media's role in polarization, assesses the effects of severe divisiveness, and recommends steps the government and the social media industry can take to ameliorate the problem. We conclude that Facebook, Twitter, and YouTube are not the original or main cause of rising U.S. political polarization, a phenomenon that long predates the social media industry. But use of those platforms intensifies divisiveness and thus contributes to its corrosive consequences. This conclusion is bolstered by a close reading of the social science literature, interviews with sociologists and political scientists who have published studies in this area, and Facebook's own pattern of internally researching the polarization problem and periodically adjusting its algorithms to reduce the flow of content likely to stoke political extremism and hatred.

Political polarization is a complicated concept. Democracy entails disagreement. Democrats clash with Republicans over taxes, immigration, and other issues, while demands for social justice may provoke controversy and backlash. In other words, in a democratic system, politics naturally creates some degree of polarization. But in light of the harmful consequences of the extreme divisiveness now plaguing the U.S., limiting polarization ought to be an urgent priority.

We focus on “affective polarization,” a form of partisan hostility characterized by seeing one's opponents as not only wrong on important issues, but also abhorrent, unpatriotic, and a danger to the country's future. This kind of hatred now infects American politics, and social media has helped spread the disease. But as we illustrate, affective polarization and its consequences are not distributed evenly across the political spectrum. Donald Trump's presidency and his continued influence over many conservatives have helped push the right to further extremes than the left has gone. January 6 provides a vivid example.

Our recommendations for diminishing the degree to which social media heightens affective polarization reflect several themes: This phenomenon constitutes a continuing threat to our democracy and requires strong responses from President Biden, Congress, and the social media industry itself. Ideally, the major platforms would have addressed these problems themselves. But having failed to self-regulate sufficiently, the companies have created a need for Washington to intervene. Only by means of vastly more disclosure about how their algorithms rank, recommend, and remove content will the platforms be held accountable for the damage they now cause to the political system and society at large.

Here, in capsule form, are our recommendations:

Recommendations in Brief

To the federal government:

1 **President Biden needs to prioritize a broad government response to the heightening of partisan hatred by social media.** By means of one or more speeches, a bipartisan blue-ribbon commission, or via some other high-visibility vehicle, Biden should seek to persuade both lawmakers and the public that to avoid future versions of the Capitol insurrection, we must confront online polarization and its malign consequences.

2 **The House Select Committee investigating the Capitol insurrection should devote ample resources to determining how technology was used to incite the violence on January 6.** Panel members must make this a central line of inquiry and use their subpoena power to pursue it.

3 **Lawmakers ought to pass legislation mandating more disclosure about the inner workings of social media platforms.** This transparency will allow outside researchers to study how algorithms decide who sees what content so that policy makers, in turn, can craft more informed legislation addressing the pathologies associated with social media use.

4 **Congress should empower the Federal Trade Commission to draft and enforce new standards for industry conduct.** Greater transparency is necessary but not sufficient. We advocate legislation authorizing the FTC to collaborate with social media companies and other stakeholders to create industry standards that would be enforceable by the government.

5 **While they grapple with social media as it now exists, legislators need to encourage exploration of alternatives to current business models.** Some technologists and entrepreneurs are imagining a radically different, pro-democratic digital future; they deserve public support.

To the platforms:

6 **Social media companies should adjust algorithms to depolarize platforms more systematically.** The platforms should create metrics to measure polarization and improve the “dial-turning” measures they now apply on an ad hoc basis to reduce antagonism during emergencies.

7 **But depolarization must take place transparently.** Disclosing what they’re doing, how they’re doing it, and what content might potentially get blocked in the process is the only way the platforms can counter suspicions that such measures are designed to manipulate politics or otherwise exert illegitimate influence.

8 **Facebook, Twitter, and YouTube should each double the size of their human content-moderation corps and make moderators full-fledged employees.** This expansion would be expensive, but it would afford front-line reviewers more time to consider difficult content decisions. Bringing them in-house would lead to better supervision of reviewers and more careful analysis.

9 **The industry needs to strengthen engagement with civil society groups that can help identify sources of dis- and misinformation related to elections, public health, and patterns of discrimination.** Social media companies should do much more to aid the growing number of nonprofits, including introducing new ways for them to share information with the platforms and one another.

10 **The platforms should reduce rewards for virality, which can contribute to polarization.** Obscuring “like” and share counts, for example, might encourage consideration of content on its merits, rather than on whether it provokes outrage, hatred, or fear.

Introduction

“Contrary to Facebook’s contentions, a range of experts have concluded that social media does contribute to polarization. Writing in *Science*, a group of 15 university researchers concluded: ‘In recent years, social media companies like Facebook and Twitter have played an influential role in political discourse, intensifying political sectarianism.’”

In March 2021, Facebook executives circulated a memo to employees seeking to knock down the idea that the company’s social media platforms contribute to political polarization. The communication arrived just two months after the insurrection at the U.S. Capitol, a shocking event fueled by false claims of a “stolen election” spread on Facebook, Twitter, and other platforms.

Troubled that the company was being implicated in toxic political developments, Facebook’s leadership used the employee memo to address what it called “an albatross public narrative for the company.” According to that narrative, “Facebook is contributing to a social problem of driving societies into contexts where they can’t trust each other, can’t share common ground, can’t have conversation about issues, and can’t share a common view on reality.” But this portrayal, management asserted, simply isn’t true: “The media narrative in this case is generally not supported by the research.”[1]

Facebook has made similar disavowals in public statements. “Some people say that the problem is that social networks are polarizing us, but that’s not at all clear from the evidence or research,” the company’s founder and chief executive, Mark Zuckerberg, testified before a U.S. House of Representatives subcommittee in March 2021. He pointed to alternative culprits: “I believe that the division we see today is primarily the result of a political and media environment that drives Americans apart.”[2] A few days later, Nick Clegg, Facebook’s vice president for global affairs and communication, argued in an article on Medium: “What evidence there is simply does not support the idea that social media, or the filter bubbles it supposedly creates, are the unambiguous driver of polarization that many assert.”[3]

Facebook’s damage-control efforts respond to conventional wisdom in Washington, D.C., where members of Congress and expert witnesses testifying before congressional committees have cited polarization as a reason to rein in the social media industry. Referring to major tech platforms, Tristan Harris, a former design ethicist at Google, told a Senate panel in April 2021: “Their business model is to create a society that is addicted, outraged, polarized, performative, and disinformed.”[4] For their part, mainstream media outlets treat social media’s role in fostering polarization as an established fact. *The New York Times* has labeled Facebook “one of the world’s most polarizing corporations,” whose “business model is optimized to keep people scrolling their Facebook feeds, amplifying divisive and inflammatory content and exaggerating political divisions in society.”[5]

“Acknowledging that, in some contexts, polarization is unavoidable doesn’t mean we should be indifferent to social media’s effect on partisan hatred more generally.”

So, who’s right? Do Facebook, and other tech platforms like Twitter and YouTube (which is owned by Google), contribute to the pernicious level of divisiveness that currently characterizes American politics? What does social science research, in fact, show? And if there is a connection, what should be done about it? These questions are urgent now and will grow more pressing as the country turns its attention to elections in 2022 and beyond.

For starters, social media is not the original or main cause of rising political polarization in America. Polarization began increasing decades before the invention of social media. Other forces —including political party realignment, hyper-partisan talk radio and cable TV, and the uniquely divisive presidency of Donald Trump—have all heightened divisiveness (see sidebar on page 8).

But that doesn’t exonerate the major social media companies. Contrary to Facebook’s contentions, a range of experts have concluded that use of social media does contribute to partisan animosity in the U.S. A group of 15 researchers offered a nuanced assessment in an article published in October 2020 in *Science*. “In recent years,” they wrote, “social media companies like Facebook and Twitter have played an influential role in political discourse, intensifying political sectarianism.” Reinforcing the point, a separate quintet of researchers summed up their review of the empirical evidence in an August 2021 article in *Trends in Cognitive Sciences*: “Although social media is unlikely to be the main driver of polarization,” they concluded, “we posit that it is often a key facilitator.”[6] In other words, in a society that has become increasingly polarized, use of social media may not *create* partisan divisiveness, but it does *exacerbate* it.

Before the extraordinary violence at the Capitol on January 6, Facebook itself had acknowledged a link between social media and polarization. In May 2020, the company posted an article on its corporate blog entitled, “Investments to Fight Polarization.” Written by Guy Rosen, vice president for integrity, the post pointed to “some of the initiatives we’ve made over the past three years to address factors that can contribute to polarization.” The initiatives included hiring more moderators to remove incendiary content, combating hate speech more aggressively, and adjusting users’ News Feeds to prioritize posts by friends and family over those of news publishers.[7] Presumably, the company wouldn’t have thought it necessary to reduce polarizing aspects of its platform if those aspects didn’t exist in the first place. In a written statement, Facebook says: “Several studies show that social media is not the primary driver of harmful polarization. But we still have a role to play in addressing it, so we remove harmful content, limit misinformation, and connect people with reliable information.”

Different types of polarization

Polarization is a complicated concept. At its worst, severe partisan alienation can undermine faith in elections and democracy itself.[8] On the other hand, in a two-party system, it’s natural for Democrats and Republicans to disagree sharply over important issues like taxation, aid to the poor, or immigration. A reasonable degree of “issue polarization,” at least in theory, should provide voters with a healthy range of policy proposals to choose from.

In this report, we primarily consider the relationship between social media and “affective polarization.” Affective polarization concerns the degree to which political opponents regard their foes as abhorrent and irredeemable. This kind of severe partisan hatred precludes political compromise, as foes come to see the other side as an existential threat to democracy. Academics measure affective polarization on a “feeling thermometer” scale ranging from cold (0°) to neutral (50°) to warm (100°). Attitudes toward opposing partisans in the U.S. plummeted from 48° in the 1970s to 20° in 2020.[9] Jonathan Haidt, a social psychologist at NYU's Stern School of Business, flips the temperature-related metaphor: While multiple forces have caused affective polarization, in recent years, social media "has become a powerful accelerant for anyone who wants to start a fire."[10]

Throughout his presidency, Donald Trump demonstrated how social media can be used to heighten racial animus and drive affective polarization. His rapid-fire posts on Twitter and Facebook sought to persuade supporters that malign forces had hijacked their country. Trump's targets included the mainstream media, which he labeled "the enemy of the people"; an "invasion" of Latin American asylum seekers whose ranks, he said, included gang members, rapists, and other "very bad people"; a quartet of liberal congresswomen of color, whom he accused of hating America and supporting Islamic terrorism; and the Black Lives Matter movement, which he accused of "treason" and "sedition."[11]

There are yet more wrinkles to polarization. In some contexts, increased polarization may accompany laudable, if controversial, political developments. For generations, demands for racial equality have sparked white backlash and political divisiveness. Assertions of rights for women and LGBTQ people likewise have contributed to polarization. The pursuit of social justice in the U.S. tends to engender resistance from the political right. That antipathy isn't a legitimate reason for squelching debate about race, gender roles, or sexual orientation. But acknowledging that, in some contexts, polarization is unavoidable doesn't mean we should be indifferent to social media's effect on partisan hatred more generally.

This report examines affective polarization not out of a fascination with readings from the feeling thermometer, but because of the consequences of severe divisiveness. Some of these consequences have appeared on the left, such as when certain Black Lives Matter protests during the summer of 2020 devolved into looting, arson, and assaults on police. But far more often in recent years, it has been the political right that has driven polarization and its corrosive effects. This asymmetry can be seen in the paralyzing dysfunction that grips Congress, the erosion of trust in democratic norms, and the undermining of commonly held facts. As illustrated by the January 6 attack on the Capitol—an event incited and organized on social media—political violence is now a real threat to American democracy and could resurface in the future.[12]

Social media and human rights

Since 2017, the NYU Stern Center for Business and Human Rights has published a series of reports about the social media industry. Past papers have explored the spread of foreign and domestically generated disinformation, the shortcomings of content moderation by social media companies, and the false claim that these companies systematically censor conservatives. The Center has undertaken this work because the operation of social media platforms affects core human rights, including freedom of expression and participation in free and fair elections. We aim not just to diagnose the detrimental effects the industry has on democracy, but also to make practical recommendations for how to ameliorate those effects.[13]

In Part 1, we assess the debate over the relationship between social media and affective polarization. Based on a review of more than 50 social science studies and interviews with more than 40 academics, policy experts, activists, and current and former industry people, we conclude that social media has intensified pre-existing polarization in the U.S. This conclusion is reinforced by Facebook's internal attempts to address aspects of polarization, which we also describe. While it is not necessarily more culpable than the other major platforms, Facebook is our primary focus for three reasons: It is the largest player in the industry, with nearly 2.9 billion global users of its main platform and 1 billion users of its Instagram platform. Additionally, while none of the major platforms has been particularly transparent, Facebook has provided more information for us to analyze, as compared to rivals like YouTube and Twitter, which are also part of the problem. Finally, most of the academic research on social media and polarization has examined Facebook.

In Part 2, we describe the asymmetric consequences of polarization in the contemporary United States. We also examine an argument by some scholars and activists that social media platforms should crack down on one particularly pernicious pathology—white supremacy—rather than worry about polarization more broadly.

In Part 3, we offer a series of recommendations to social media companies and lawmakers. We argue that industry and government can each take practical, if not necessarily easy, steps to reduce the overall level of polarizing content online.

Part 1: Assessing the Research on Social Media and Polarization

Researchers have studied social media's relationship to political polarization for more than a dozen years. As noted in the Introduction, Facebook argues that this research does not reveal a link between social media and growing political divisiveness. But a close look at the research, and interviews with many of the academics who have produced it, shows that there is an important connection between social media use and partisan animosity.

> "Studies of polarization that encompass decades before the advent of social media naturally tend to point to other factors as likely causes of increased divisiveness. But research focused more narrowly on the years since 2016 suggests that widespread use of the major platforms has exacerbated partisan hatred."

Matthew Gentzkow, an economics professor at Stanford, has co-authored a number of the leading empirical studies on social media and polarization, including research that Facebook cites. Gentzkow's assessment of the academic literature is markedly different from the company's, however. Looking back to the 2000s and the introduction of Facebook, Twitter, and YouTube, it's not clear "what role social media has played in the long-term, broad increase in polarization we've seen in the U.S.," he says in an interview. But more recent evidence, especially since Donald Trump's election as president in 2016, he adds, points to the conclusion that "social media can cause polarization."

In one study, Gentzkow and fellow researchers paid American subjects to stop using Facebook for a month, until just after the 2018 midterm elections. The randomized study involved 2,743 people, including a control group that continued to use Facebook. After the experiment, the researchers surveyed participants and reported their results in March 2020. Staying off Facebook, they found, "significantly reduced polarization of views on policy issues" but didn't reduce affective polarization in a statistically significant way. "That's consistent with the view that people are seeing political content on social media that does tend to make them more upset, more angry at the other side [and more likely] to have stronger views on specific issues," Gentzkow says in the interview. The study also found that taking a break from Facebook reduced participants' knowledge of news events but enhanced their subjective sense of well-being.[14]

While the Facebook "deactivation" experiment looked at recent effects in a short-term context, another study Gentzkow helped oversee assessed longer-term effects. The latter research compared increases in polarization levels between 1996 and 2016 for three age groups: 18–39, 40–64, and 65 and older. Published in September 2017, the age-based analysis found that polarization increased the most among those who used the internet and social media the least—namely, the 65-and-older group.[15] This suggests that "social media is not the main reason why polarization has been going up in the U.S. over time," Gentzkow says.

The Trump factor

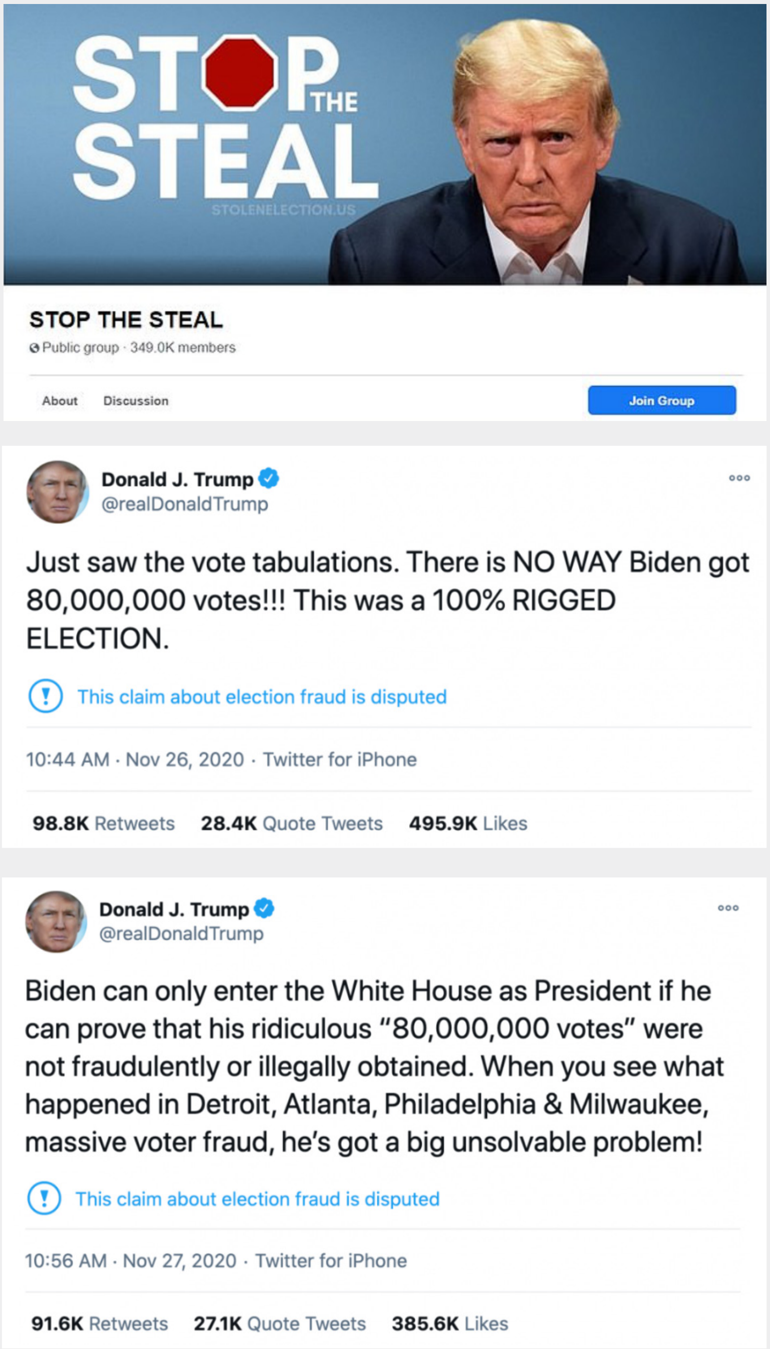
But the age-group analysis requires qualification. For one thing, today's major social media platforms weren't started until roughly half-way through the 20-year period the study covers. Even more significantly, the study ended in 2016, a year that marked a turning point in the evolution of social media as it relates to polarization. Even before he took office, Trump and his supporters pursued an unprecedented social media campaign aimed at provoking us-versus-them hatred in America.[16] The Trump factor, however, wasn't reflected in the age-group study. Gentzkow says that it's possible that events since 2016, including Trump's campaign to undermine the 2020 election, have made "the forces by which social media can drive polarization stronger than they've been in the past."

A third study Gentzkow worked on compared polarization rates in the U.S. and eight other developed democracies over four decades. Published in January 2020, the paper found that the U.S. experienced the largest increase in affective polarization. In three other countries—Canada, New Zealand, and Switzerland—polarization also rose, but to a lesser extent. In five countries—Australia, Britain, Germany, Norway, and Sweden—polarization fell. Given the global nature of the internet, and now, social media, these varying results don't suggest that social media has driven the long-term increase in polarization in America, the authors concluded.[17] Facebook has drawn attention to the inter-country comparison, as have certain prominent analysts. Columnist and podcaster Ezra Klein of *The New York Times*, who wrote the 2020 book *Why We're Polarized*, has asserted that the inter-country study "lets us reject [the idea] that polarization is a byproduct of internet penetration or digital media usage."[18]

Both Facebook and Klein go too far in interpreting this research. Granted,

Divisive content

On November 4 and 5, 2020, several hundred thousand Trump supporters joined Stop the Steal, a Facebook Group (top) committed to the falsehood that the election had been stolen from Trump. Although Facebook removed the Group, Trump continued to promote the lie – for example, on Twitter (middle, bottom) – until both platforms barred his accounts in January 2021.



STOP THE STEAL
Public group · 349.0K members

Donald J. Trump
@realDonaldTrump

Just saw the vote tabulations. There is NO WAY Biden got 80,000,000 votes!!! This was a 100% RIGGED ELECTION.

This claim about election fraud is disputed

10:44 AM · Nov 26, 2020 · Twitter for iPhone

98.8K Retweets 28.4K Quote Tweets 495.9K Likes

Donald J. Trump
@realDonaldTrump

Biden can only enter the White House as President if he can prove that his ridiculous "80,000,000 votes" were not fraudulently or illegally obtained. When you see what happened in Detroit, Atlanta, Philadelphia & Milwaukee, massive voter fraud, he's got a big unsolvable problem!

This claim about election fraud is disputed

10:56 AM · Nov 27, 2020 · Twitter for iPhone

91.6K Retweets 27.1K Quote Tweets 385.6K Likes the study highlights factors other than social media that have contributed to polarization in the U.S. But the more important insight it reveals is that the causes and accelerants of polarization vary from nation to nation and probably from time period to time period. As co-author Gentzkow notes, his analysis of decades-long trends up to 2016 does not speak directly to the drivers of affective polarization since the watershed 2016 election.

Maximizing user engagement

The 15-scholar collective that published the 2020 *Science* article on political sectarianism added an important element to this discussion. They identified the feature of social media that is primarily responsible for the amplification of divisive content. This feature is the fundamental design of the automated systems that run the platforms. “Social media technology employs popularity-based algorithms that tailor content to maximize user engagement,” the co-authors wrote. Maximizing engagement increases affective polarization, they added, especially within “homogeneous networks,” or groupings of like-thinking users. This is “in part because of the contagious power of content that elicits sectarian fear or indignation.”[19]

Social media companies do not seek to boost user engagement because they want to intensify polarization. They do so because the amount of time users spend on a platform liking, sharing, and retweeting is also the amount of time they spend looking at the paid advertising that makes the major platforms so lucrative. Content that elicits partisan fear or indignation is particularly contagious and helps fuel this advertising business model. In 2020, advertising provided 98% of Facebook’s $86 billion in revenue. Google, which includes YouTube, reported $182 billion in revenue, 81% of which came from advertising.[20] Facebook, in a written statement, counters: “We’ve long said that it is not in our interest—financially or reputationally—to turn up the temperature or push users towards ever more extreme content.”

In her 2018 book, *Frenemies: How Social Media Polarizes America*, Jaime Settle argued that a number of other Facebook features heighten divisiveness among users. An associate professor of government at the College of William & Mary, Settle observed that while a relatively small percentage of people set out to find political content online, the “vast majority of Facebook users are potentially ‘dosed’ with polarizing informative content.” Having surveyed more than 3,000 people, she found that Facebook makes it easy to infer other users’ partisan views and encourages a fusion of political and social identities. This, in turn, fosters stereotyped evaluations of those with whom users disagree and politicizes non-political issues.[21] Illustrating the latter

Race, Realignment, and Cable TV

Polarization hit a post-Civil War low in the middle of the 20th century, when the Democratic and Republican Parties each represented broad, politically heterogeneous coalitions of liberals and conservatives. In fact, the American Political Science Association issued a report in 1950 calling for *more* polarization, saying that voters deserved more distinct policy choices.[1]

Over the following decades, the political scientists got what they asked for—and then some. Social conflict largely related to race prompted a resorting of the political parties. Repelled by calls for Black voting rights and racial integration, many southern Democrats became Republicans. A number of liberal Republicans switched parties, while others were defeated at the polls. The parties became more distinct when measured along urban/rural, educational, and religious lines. This resulted in a more liberal Democratic Party and an increasingly conservative Republican Party.

Other factors also contributed to polarization. By the 1990s, the “big three” broadcast news networks, which prized their reputation for impartiality, had receded in influence. Conservative talk radio and Fox News provided Republicans with an unabashedly right-wing version of political events. On the left, MSNBC, later joined by CNN, countered with a liberal take on the news, although one generally less extreme than that of their counterparts on the right.

Political leaders also moved further apart in ideological terms, but again, the trend was asymmetrical, with Republicans migrating further to the right than the Democrats moved to the left.[2] By the 2010s, the flames of affective polarization were already intense; social media provided a ready accelerant.

[1] https://www.jstor.org/stable/i333592

[2] https://science.sciencemag.org/content/370/6516/533

phenomenon, many people on the political right who protested pandemic stay-at-home orders, maskmandates, and vaccination programs said they did so because the public health measures trampled on their personal freedom or were part of a government conspiracy.[22] In an interview, Settle emphasizes that social media is one of several factors that feed polarization and threaten to erode “the glue that holds society together.”

‘Echo chambers’

A subpart of the polarization debate concerns the question of whether social media fosters “echo chambers,” also known as “information cocoons,” in which partisans hear only one side of the story and develop animosity toward anyone who believes the other side. Cass Sunstein, a professor at Harvard Law School now working in the Biden Administration, has sounded the alarm about this danger. “If you live in an information cocoon,” he has written, “you will believe many things that are false, and you will fail to learn countless things that are true. That’s awful for democracy.” According to Sunstein, when social media users encounter unexpected or opposing viewpoints, democratic discourse benefits.[23]

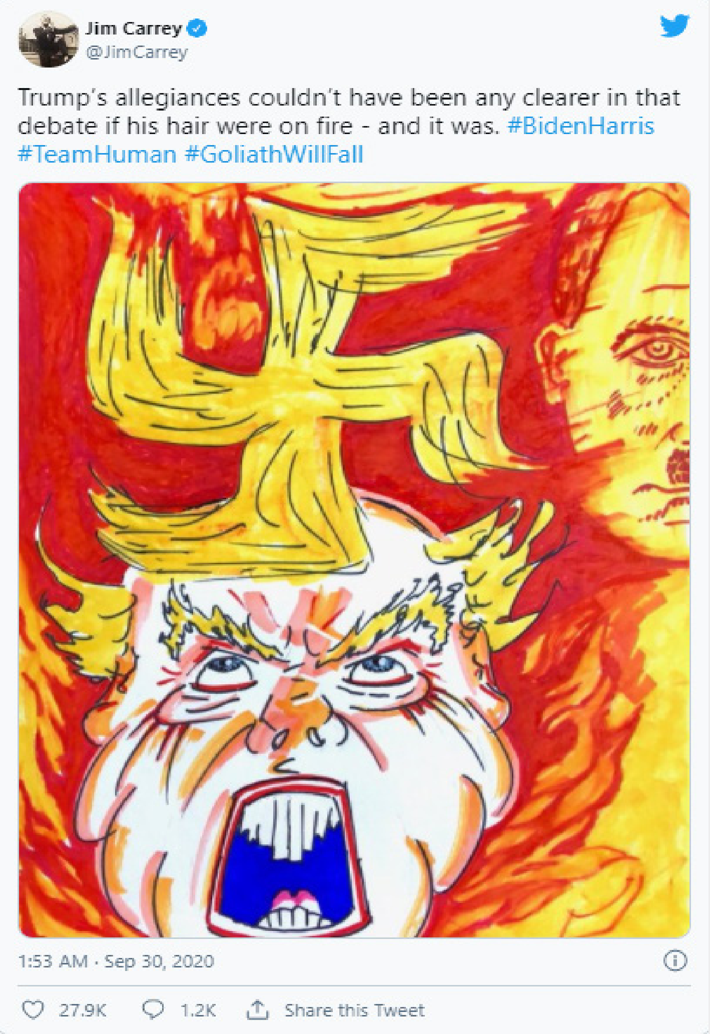
Ro’ee Levy of the Tel Aviv University School of Economics made a similar finding in a paper published in March 2021. Based on a study of more than 17,000 American participants, Levy found that Facebook’s content-ranking algorithm may limit users’ exposure to news outlets offering viewpoints contrary to their own—and thereby increase polarization.[24] Research by Facebook itself, published in 2015, reached a different conclusion, finding that user behavior on the platform plays a larger role than algorithmic ranking in limiting exposure to contrary content.[25]

Others have suggested that exposure to opposing ideologies on social media can push users to greater extremes. Researchers led by Christopher Bail, a sociology professor at Duke University, conducted a randomized experiment in which roughly half of the 1,220 participants received financial incentives to follow a Twitter bot for a month that exposed them to opposing ideologies as expressed by elected officials, opinion leaders, and media organizations. In a paper published in September 2018, the researchers reported that Republicans who followed a liberal bot “became substantially more conservative.”

In contrast, Democrats who followed a conservative bot exhibited increases in liberal attitudes so slight they weren’t statistically significant. The researchers drew a sobering conclusion, one that contradicts Sunstein’s and Levy’s work: “Attempts to introduce people to a broad range of opposing political views on a social media site such as Twitter might be not only ineffective but counterproductive.”[26]

Divisive content

In September 2020, comic actor Jim Carrey, who has 18.6 million followers on Twitter, depicted a raging then-President Trump beneath a Nazi swastika.



Jim Carrey
@JimCarrey

Trump’s allegiances couldn’t have been any clearer in that debate if his hair were on fire - and it was. #BidenHarris #TeamHuman #GoliathWillFall

1:53 AM · Sep 30, 2020

Expanding on the 2018 echo chamber study, Bail published a book in 2021 entitled, *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. It assesses thousands of social media users based on a combination of hundreds of millions of data points and in-depth interviews. Bail came to a two-part conclusion: First, "the root source of political tribalism on social media lies deep inside ourselves," tapping fears and resentments about race, money, privilege, religion, and more. But the medium itself distorts and amplifies these already-roiling emotions: "The social media prism fuels status-seeking extremists, mutes moderates who think there is little to be gained by discussing politics on social media, and leaves most of us with profound misgivings about those on the other side."[27]

YouTube and the 'rabbit hole'

The polarizing influence of social media isn't limited to Facebook. YouTube has been singled out by academics, journalists, and others for radicalizing some users by guiding them to ever-more-extreme videos. Seventy percent of the time spent on YouTube stems from suggestions by the platform's recommendation algorithm.

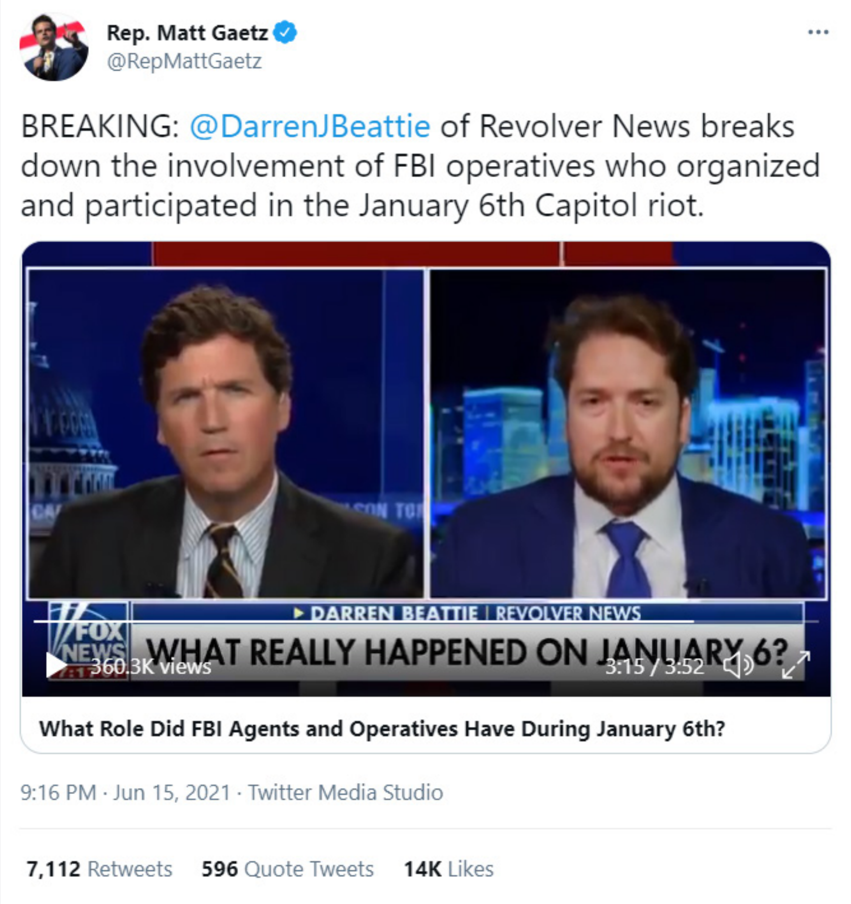
Zeynep Tufekci, a sociologist and associate professor at the University of North Carolina, has written that if a user indicates an interest in either left- or right-leaning politics, the YouTube algorithm will recommend that they sample more provocative and divisive fare. Describing her own experience in 2018, Tufekci noted that when she clicked on videos of Trump rallies, "YouTube started to recommend and 'autoplay' videos for me that featured white supremacist rants, Holocaust denials, and other disturbing content." After creating a fresh account and seeking out videos of Hillary Clinton and Bernie Sanders, Tufekci received recommendations involving "secret government agencies and allegations that the United States government was behind the attacks of September 11." YouTube, she concluded, "leads viewers down a rabbit hole of extremism."[28]

Researcher Jonathan Albright has described something similar. Formerly the director of the digital forensics initiative at the Tow Center at Columbia Journalism School, Albright wrote in 2018 about doing a YouTube search for "crisis actor," a term used by some on the far right to describe students who they falsely claim have taken part in staged school massacres to encourage tougher gun control. The crisis actor videos, according to Albright, led to recommendations featuring "celebrity pedophilia, 'false flag' rants, and terror-related conspiracy theories dating back to the Oklahoma City attack in 1995."[29]

YouTube says that it has made changes to address the rabbit-hole phenomenon: "Over the past few years," the platform explains in a written statement, "we've invested heavily in the policies, resources, and products needed to protect the YouTube community. We changed our search and discovery algorithms to ensure more authoritative content is surfaced and labeled prominently in search results

Divisive content

In June 2021, Representative Matt Gaetz (R.,Fla.) used Twitter to amplify the false inside-job theory, earlier promoted by Tucker Carlson of Fox News, that the FBI instigated the Capitol riot.



Rep. Matt Gaetz
@RepMattGaetz

BREAKING: @DarrenJBeattie of Revolver News breaks down the involvement of FBI operatives who organized and participated in the January 6th Capitol riot.

FOX NEWS
DARREN BEATTIE | REVOLVER NEWS
WHAT REALLY HAPPENED ON JANUARY 6?
360.3K views
3:15 / 3:52

What Role Did FBI Agents and Operatives Have During January 6th?

9:16 PM · Jun 15, 2021 · Twitter Media Studio

7,112 Retweets **596** Quote Tweets **14K** Likes and recommendations, and began reducing recommendations of borderline content that could misinform users in harmful ways. In the year following this change, watch time of borderline content from non-subscribed recommendations dropped by over 70% in the U.S, and we saw a similar drop in other markets as well."

YouTube hasn't released the underlying data necessary to verify the claimed 70% drop. But recent research supports the company's contention that, if its recommendation function once pushed some users to political extremes, that tendency has been significantly reduced. A study publishedin August 2021 in *Proceedings of the National Academy of Sciences* assessed the individual-level browsing behavior of more than 300,000 American YouTube users from January 2016 through December 2019. The six co-authors said they found "no evidence that engagement with far-right content is caused by YouTube recommendations systematically." Rather, they added, "consumption of political content on YouTube appears to reflect individual preferences that extend across the web as a whole."[30]

Users, of course, may search for potentially harmful content on YouTube without help from the platform's recommendation algorithm. To evaluate how often users encounter videos that violate the site's guidelines because they contain violence, hate speech, harassment, or the like, YouTube has begun publicly reporting its "violative view rate," or VVR. In April 2021, YouTube said the VVR stood at 16 to 18 instances of violative content per 10,000 views. This reflected a 70% decrease since 2017, the video platform said.[31]

In thinking about the violative view rate, it's important to keep in mind YouTube's heft—specifically, how many views its users actually tally in a typical day. The platform has more than two billion monthly users worldwide, and cumulatively, they watch about a billion hours of YouTube videos a day, according to the company. This activity generates billions of individual views on a daily basis.[32] YouTube doesn't specify how many billions of daily views its users tally. But at that scale, the number of daily views of harmful content would be measured in the millions—not an insubstantial amount.

'Turning the dial' at Facebook

The main locus of research on social media and polarization is not at one or another major university but at Facebook. The company employs hundreds of social scientists to study various aspects of Facebook's effects on users and society at large. Little of their work sees the light of day, but its very existence suggests that the company is concerned about how social media use affects democracy. That Facebook invests in such extensive self-analysis is laudable, but the introspection on polarization probably would be more productive if the company's top executives were not publicly casting doubt on whether there is any connection between social media and political divisiveness.

Facebook relies on its in-house research as a basis for periodic adjustments to algorithms for ranking, recommending, and removing content, according to Yann LeCun, who is both the company's vice president and chief scientist for artificial intelligence and a professor at NYU's Courant Institute of Mathematical Sciences. Political polarization "has been an issue that's front and center for a lot of people at Facebook," he says in an interview, "but it's difficult to correct for it, so the things that are done are not widely advertised." (LeCun notes that he does not have his comments vetted by Facebook.)

Now and then, the company does go public about steps it has taken to mitigate polarization. And on other occasions, journalists have unearthed evidence of Facebook's consideration of emergency measures meant to limit the level of extreme and divisive content. These measures are sometimes referred to as "turning the dial" or "breaking glass," as in breaking the glass of a fire alarm box.

In May 2020, *The Wall Street Journal* reported on an internal Facebook analysis that concluded that its automated ranking and recommendation systems were driving people apart. "Our algorithms exploit the human brain's attraction to divisiveness," a slide from a presentation in 2018 stated. Without changes, the presentation added, the company's automated systems would steer users to "more and more divisive content in an effort to gain user attention & increase time on the platform." Despite the warning, CEO Mark Zuckerberg and other senior executives "largely shelved the basic research," the *Journal* reported. One reason for the decision not to make changes was a fear that adjustments to dampen divisiveness would disproportionately crimp conservatives, some of whom were already alleging that Facebook censored them.[33]

Facebook responded to the *Journal* article with the corporate blog post "Investments to Fight Polarization," which we mentioned in the Introduction. The post accused the news organization of placing too much emphasis on "a couple of isolated initiatives we decided against." The company pointed to four steps it had taken over the previous several years to lessen divisiveness: First, it altered users' News Feeds to prioritize posts from friends and family over news content. "This was based on extensive research that found people derive more meaningful conversations and experiences when they engage with people they know rather than passively consuming content," Facebook said. Second, it "built a global team of more than 35,000 people working across the company on issues to secure the safety

> ‘Our algorithms exploit the human brain’s attraction to divisiveness,’ a slide from a 2018 internal Facebook presentation stated, according to *The Wall Street Journal*. Without changes, the presentation added, the company’s automated systems would steer users to ‘more and more divisive content in an effort to gain user attention and increase time on the platform.’

and security of our services, including those related to polarization.” About 15,000 of those people are front-line content moderators, the vast majority of whom are employed by third-party outsourcing firms, not by Facebook. Third, the company restricted recommendations to, and content spread by, Pages and Groups that repeatedly violated the platform’s rules or were deemed to have distributed falsehoods. Finally, the company expanded “pro-active detection technology” to remove hate speech more quickly.[34]

Facebook has taken these and other actions even in the absence of what the company considers definitive evidence that its platforms contribute to political polarization, a Facebook official says in an interview, adding that it does so out of a sense of responsibility to its users and because the actions are salutary in their own right. Toward the same end, the platform has pointed users to “information centers” on the site where they can find authoritative material about potentially polarizing topics such as Covid-19, elections, and climate change.

“This is not to say that the measures that have been taken to counteract [divisiveness and extremism] have been perfectly successful or efficient, because it’s really hard,” LeCun, the company’s AI chief, says in a separate interview. On Twitter, LeCun has argued that social media shouldn’t be blamed for causing polarization in the first instance. But the major platforms, he says in the interview, “could have helped accelerate the process.”[35]

‘News ecosystem quality’

Anticipating an ugly, disinformation-drenched 2020 presidential campaign, Facebook prepared a battery of dial-turning measures. Ahead of Election Day, it temporarily altered its recommendation algorithm to stop steering users to politically oriented Groups, some of which had become hotbeds of right-wing extremism and violent talk. While well intended, the pause on recommendations didn’t prevent Group members from inviting new recruits to join; nor did it prevent users from taking the initiative to search for Groups whose members were spreading the false idea that the election would be rigged against Donald Trump.[36]

When, in the aftermath of his defeat, Trump and his supporters took to Facebook, Twitter, and other platforms to spread falsehoods about rogue ballots and fixed voting machines, certain Facebook employees pushed for additional measures meant to calm the tone and content of users’ News Feeds. According to *The New York Times*, Zuckerberg agreed, and more weight was applied to what the company calls “news ecosystem quality,” or NEQ. A secret internal ranking assigned to publishers based on signals of their authoritativeness, NEQ favors mainstream organizations like the *Times* and NPR and disfavors hyper-partisan outlets like Breitbart on the right and Occupy Democrats on the left.[37]

Some Facebook employees argued that the NEQ tweak had created a “nicer News Feed,” one that should continue even after the contentious post-election period. But that didn’t happen. On a conference call with journalists in mid-November 2020, Guy Rosen, vice president for integrity, explained that emergency election-related changes were always intended to be temporary. “There has never been a plan to make these permanent,” he said.[38]

A number of observers have attributed Facebook’s ambivalence about tamping down extreme and polarizing content to its top management’s countervailing interest in the growth of its user base and bottom line. As a result of the algorithmic dial-turning in November 2020, “everybody saw a reduction in election-related misinformation on Facebook,” Hany Farid, a computer science professor at the University of

California, Berkeley says in an interview. But in their book, *An Ugly Truth: Facebook's Battle for Domination*, which was published in July 2021, *New York Times* journalists Sheera Frenkel and Cecilia Kang reported that company executives worried that maintaining a "nicer News Feed" beyond the election period would result in "users spending less time on the platform."[39]

Facebook has made seemingly contradictory moves with regard to outside researchers studying the platform's effect on elections. It gave a group of scholars access to certain internal data on the 2020 presidential race from which they are expected to publish studies in 2022. But separately, Facebook cut off researchers with NYU's Cybersecurity for Democracy, who were seeking to determine whether the platform has been used to sow distrust in elections, among other questions. The company accused the NYU team of gathering information improperly—an accusation the group, and a variety of expert observers, have denied.[40]

'Emergency situations'

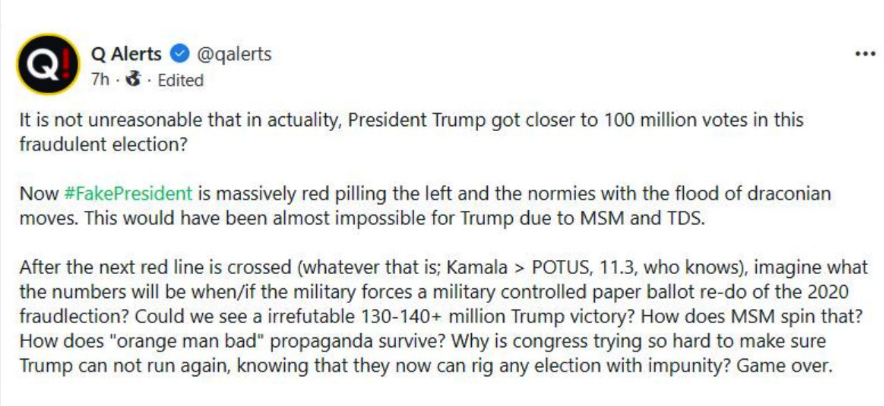
Facebook periodically continues to adjust its algorithms in attempts to reduce divisiveness and avoid real-world harm, the company official says in an interview. When political clashes in the U.S. or other countries threaten to become violent, for example, the company may temporarily instruct its automated moderation system to reduce the distribution of content where the system is relatively confident that the content violates Facebook's standards, but less confident than it would ordinarily need to be for removal. For example, if the system is 75% confident that the content is hate speech that violates Facebook's standards, its distribution would be reduced, but the content would not be removed. The reduction in distribution would be proportional to the confidence of the moderation system. Once company officials determine that the exigent circumstances have ended, the algorithms are often reset to identify and remove only content deemed with near-certainty to violate the company's standards.

This complicated process played out in April 2021, as the nation awaited a verdict in the trial of Derek Chauvin, the former Minneapolis policeman charged with murdering George Floyd. The day before Chauvin's conviction, Facebook announced that it was prepared to limit incendiary fallout online. "As we have done in emergency situations in the past," Monika Bickert, the company's vice president for content, said in a corporate blog post, "we may also limit the spread of content that our systems predict is likely to violate our community standards in the areas of hate speech, graphic violence, and violence and incitement." The Bickert post thus confirmed that, not just in connection with the Chauvin trial, but in an unspecified number of "emergency situations in the past," Facebook had turned the dial to curb the reach of toxic content.[41]

But why only in emergencies? Writing in response to Bickert, Evelyn Douek, a researcher affiliated with Harvard's Berkman Klein Center for Internet & Society, observed: "What the company

Divisive content

Until they attempted to purge their sites of QAnon content after the January 6 insurrection, Facebook and Twitter facilitated the spread of the pro-Trump conspiracy theory. Facebook even recommended QAnon groups to some users (right), while Twitter gave them a platform to press baseless claims of election fraud (below).

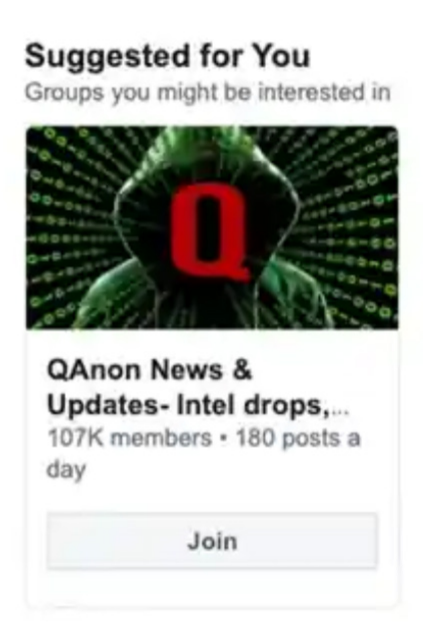


Q Alerts @qalerts
7h · Edited

It is not unreasonable that in actuality, President Trump got closer to 100 million votes in this fraudulent election?

Now #FakePresident is massively red pilling the left and the normies with the flood of draconian moves. This would have been almost impossible for Trump due to MSM and TDS.

After the next red line is crossed (whatever that is; Kamala > POTUS, 11.3, who knows), imagine what the numbers will be when/if the military forces a military controlled paper ballot re-do of the 2020 fraudlection? Could we see a irrefutable 130-140+ million Trump victory? How does MSM spin that? How does "orange man bad" propaganda survive? Why is congress trying so hard to make sure Trump can not run again, knowing that they now can rig any election with impunity? Game over.



Suggested for You

Groups you might be interested in

QAnon News & Updates- Intel drops,...

107K members • 180 posts a day

hasn't explained is why its anti-toxicity measures need to be exceptional at all. If there's a reason turning down the dials on likely hate speech and incitement to violence all the time is a bad idea, I don't see it."[42]

Subsequently, Bickert faced this question—why not permanently dial down hate speech and incitement?—during a Senate hearing where she was a witness. Her answer: Facebook sees a free-speech "cost" in doing so. The cost is that its content-removal algorithms are prone to taking down "false positives"—borderline content that, upon human inspection, turns out not to violate Facebook's standards. "It's always this balance between trying to stop abuse and trying to make sure that we're providing a space for freedom of expression and being very fair," she added.[43]

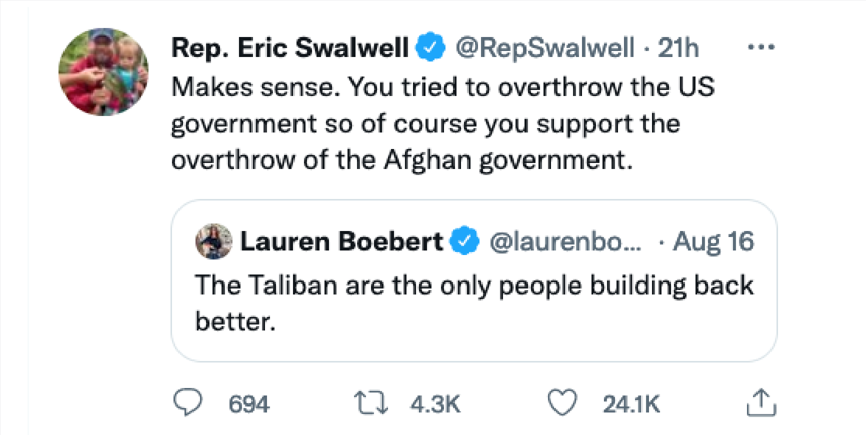
No one would argue against fairness, of course. As we discuss in our recommendations in Part 3, Facebook and other platforms can, and should, redouble their efforts to avoid removal of false positives. This entails continuing to refine content moderation algorithms while simultaneously enlarging and better training the workforce of human content moderators.

A number of former Facebook executives and employees have said that despite public declarations about stopping abuse and promoting free speech, the company's top leadership actually is more focused on putting out public relations fires. Brian Boland quit Facebook in November 2020 after working there for more than 11 years. In interviews in the summer of 2021, the former vice president for partnerships strategy explained that he left, in part, because of his growing frustration over the company's resistance to grappling with phenomena like the viral distribution of misinformation about vaccines, an issue that has contributed heavily to political polarization. The concern of senior management, Boland told CNN, "seems to be more about 'let's avoid the story' or 'let's control the narrative,' rather than 'let's do the hard thing.'"[44] Facebook declined to comment.

> "Facebook's senior management 'seems to be more about "let's avoid the story" or "let's control the narrative," rather than "let's do the hard thing."'—Brian Boland, former Facebook vice president for partnerships strategy"

Divisive content

Responding on Twitter to Representative Lauren Boebert's provocation about the Taliban's takeover of Afghanistan in August 2021, liberal Representative Eric Swalwell (D., Calif.) upped the ante by accusing the conservative Colorado Republican of having tried to overthrow the U.S. government.



Rep. Eric Swalwell @RepSwalwell · 21h ···

Makes sense. You tried to overthrow the US government so of course you support the overthrow of the Afghan government.

Lauren Boebert @laurenbo... · Aug 16

The Taliban are the only people building back better.

694 4.3K 24.1K

Part 2: Asymmetric Polarization and its Consequences

In 2012, prominent congressional analysts Thomas Mann and Norman Ornstein published a book about growing political polarization in the U.S. called, *It's Even Worse Than It Looks*. Mann, affiliated with the liberal Brookings Institution, and Ornstein, with the conservative American Enterprise Institute, argued that Democratic and Republican lawmakers had moved apart over previous decades, but Republicans had moved much further to the right than Democrats had to the left. This asymmetric polarization, the authors concluded, had paralyzed Congress and threatened to undermine our constitutional democracy.[45]

In the nearly ten years since the appearance of the Mann-Ornstein book, as social media has helped exacerbate partisan hatred, the asymmetry of political polarization in the U.S. has grown even more acute. The severity of the situation appeared in stark relief during the January 6 pro-Trump insurrection, to name only the most dramatic example. Five months later, a May 2021 *Economist*/YouGov poll found that 22% of Republicans have at least a somewhat favorable opinion of the people who stormed the Capitol.[46] That troubling frame of mind has persisted. In July 2021, a CBS News/YouGov poll found that 25% of Republicans approved of the rioters.[47]

Having established in Part 1 that social media has heightened partisan hatred in the U.S., we turn now to the consequences of this relationship, which dominate the nation's politics across a number of dimensions: Trust among citizens and in important institutions has declined. Among Trump supporters, allegiance to democratic norms, like elections and the peaceful transfer of power, has eroded. Many of the same people reject objectively grounded facts and instead promote rank falsehoods—for example, about whether Covid-19 vaccines are safe and effective. For their part, members of a dysfunctional Congress cannot find common ground to address climate change, gun violence, and other dire problems. And ultimately, the most extreme political alienation has led to threats, intimidation, and violence.

> “The consequences of polarization now dominate the nation's politics across a number of dimensions: the decline in trust of fellow citizens and important institutions; the rejection of shared facts and promotion of falsehoods; legislative dysfunction; erosion of democratic norms; and ultimately, radicalization and violent extremism.”

Decline in trust

Trust among citizens is the most basic currency of any society. For decades, as polarization has increased, trust among Americans has declined. In the early 1970s, about half of Americans said that most people can be trusted; today, that figure has fallen to less than one-third.[48]

Trust has deteriorated more on the political right than on the left. A 2019 Pew Research Center survey found that nearly two-thirds of Republicans see the other side as unpatriotic, while less than a quarter of Democrats feel that way. Asked whether the two parties are “respectful and tolerant of different

> “Vulnerability to disinformation is asymmetric, researchers from the University of Copenhagen found when they evaluated data from Twitter. Referring to pro-Russian disinformation aimed at U.S. audiences, they reported that conservative Twitter users are ‘significantly more likely to follow disinformation accounts, compared to liberal users.’”

types of people,” 60% of Pew’s respondents said this phrase described Democrats at least somewhat well, while only 38% said it described Republicans somewhat or very well.[49] An annual “trust barometer” survey of 1,500 Americans conducted by the Edelman public relations firm after the 2020 presidential election found that, compared to Biden voters, Trump voters expressed less trust in government, media, NGOs, and other institutions.[50]

QAnon, the right-wing conspiracy network that has spread via social media recommendations and group invitations, provides an extreme example of the corrosive effects of distrust. Adherents believe that a satanic cabal of high-ranking Democratic pedophiles have conspired with “deep state” bureaucrats to keep Donald Trump from his rightful place in the White House.[51] After this dark fantasy metastasized on their platforms for years, Facebook and Twitter removed tens of thousands of QAnon-related accounts in the aftermath of January 6. But zealots have learned to disguise their language as many have scattered to more obscure digital venues, seeding distrust as they go.[52]

Rejection of shared facts and promotion of falsehoods

The sense that ideological opponents now inhabit separate realities is often discussed in conjunction with the undermining of trust among Americans. Commenting on anxiety-driven purchases of firearms during the pandemic, Lilliana Mason, a political scientist at the University of Maryland who studies political violence, told *The New York Times*: “There is a breakdown in trust and a breakdown in a shared, common reality.”[53]

The rejection of basic facts related to Covid-19 correlates to polarized politics shaped in part by social media. In 2021, pro-Trump jurisdictions have seen high rates of resistance to widely available, highly effective vaccinations and corresponding high rates of infection. In areas that backed President Biden, larger percentages of residents have sought vaccination, and proportionally fewer people have contracted the coronavirus.[54]

Widespread use of social media and distrust of mainstream media appear to have combined to increase the political system’s vulnerability to partisan misinformation.[55] A study published in June 2021 by R. Kelly Garrett and Robert Bond of the Ohio State University School of Communication concluded that “U.S. conservatives are uniquely susceptible to political misperceptions in the current socio-political environment.” The research, which combined social media engagement data with a six-month longitudinal panel study of Americans’ political news knowledge, also found that “conservatism is associated with a lesser ability to distinguish between true and false claims across a wide range of political issues.” One potential explanation for the ideological discrepancy, according to the authors, is that “viral falsehoods most often promote conservative interests.”[56]

A group of Danish researchers arrived at a similar conclusion, noting in May 2021 in the Tech Stream blog of the Brookings Institution that false and manipulated facts attract more attention on the political right. Their study of Twitter data revealed “a marked lopsidedness in the political slant” of shared fake news articles: “A majority came from pro-Republican outlets and were shared by people who identified as Republicans.”[57] This observation is reinforced by conclusions reached by Henry Brady and Brad Kent of the University of California, Berkeley. Based on their analysis of polling since 1970, Brady and Kent found that conservatives more often distrust “basic knowledge-producing institutions, including higher education, science, and journalism.”[58]

Another asymmetric consequence of political polarization is vulnerability to disinformation advanced by foreign state interests. Researchers from the University of Copenhagen evaluated a large data set related to American Twitter users and found that the "reach of online, pro-Russian disinformation into U.S. audiences is distinctly ideologically asymmetric." Conservative Twitter users, they found, are "significantly more likely to follow disinformation accounts, compared to liberal users."[59]

Legislative dysfunction

Extreme polarization is a proven obstacle to legislative progress. In his 2019 book, *Polarization: What Everyone Needs to Know*, Princeton political scientist Nolan McCarty compared congressional polarization levels to lawmakers' productivity and found that the legislative branch has enacted the vast majority of its significant measures when it has been the least polarized. Specifically, the 10 least polarized congressional terms have produced roughly 16 significant enactments per term, while the 10 most polarized terms have produced slightly more than 10.[60]

In recent years, many conservatives seem to have preferred to see government programs fail rather than succeed. "An extreme form of this was on display during the Trump Administration," according to McCarty, "when the president mused aloud that it would be better to allow the Affordable Care Act (Obamacare) to implode rather than negotiate on reforms that minimize the loss of insurance coverage."

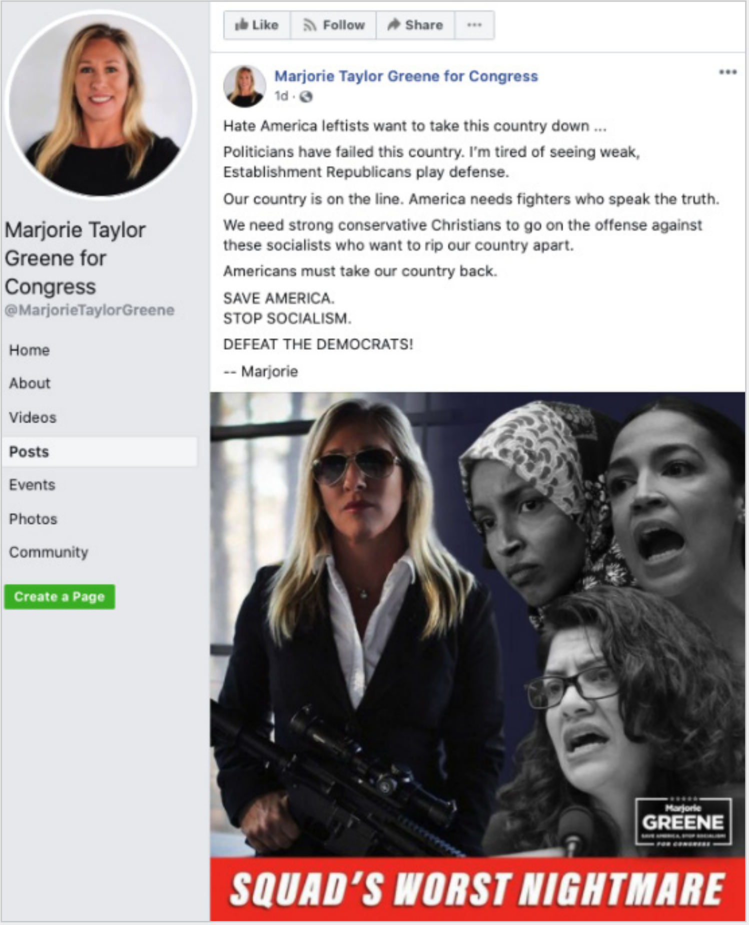
Researchers at NYU and Cambridge University recently found that the tendencies affecting social media use among ordinary voters also characterize the online activity of members of Congress. The June 2021 study analyzed Facebook and Twitter posts by both Republican and Democratic lawmakers. It concluded that "social media may be creating perverse incentives for [members of Congress to spread] divisive content because this content is particularly likely to go 'viral.'" The researchers determined that posts expressing "out-group animosity may be good at generating superficial engagement while ultimately harming individuals, political parties, or society in the long term."[61]

Erosion of democratic norms

For democracy to succeed, citizens and their political leaders must respect the outcome of elections and the peaceful transfer of power. "Without this norm of gracious losing, democracy is not sustainable," according to Harvard political scientists Steven Levitsky and Daniel Ziblatt.[62] The social media-infused affective polarization that taints U.S. politics has contributed to erosion of this foundational democratic norm. By promoting the idea that partisan foes must be kept from power at all costs, extreme us-versus-them antagonism undermines democratic values.

Divisive content

Majorie Taylor Greene demonized Democrats as "Hate America leftists" in posts such as this one on her 2020 campaign Facebook page, where she is depicted holding an AR-15-style large-capacity rifle. Since taking office as a congresswoman from Georgia, she has compared mask mandates to the requirement that Jews in Nazi Germany wear yellow stars, to name just one polarizing comment.



Hate America leftists want to take this country down ...

Politicians have failed this country. I'm tired of seeing weak, Establishment Republicans play defense.

Our country is on the line. America needs fighters who speak the truth.

We need strong conservative Christians to go on the offense against these socialists who want to rip our country apart.

Americans must take our country back.

SAVE AMERICA.
STOP SOCIALISM.

DEFEAT THE DEMOCRATS!

-- Marjorie

In a December 2018 paper, Jennifer McCoy of Georgia State University and Murat Somer of Koc University in Istanbul noted how, beginning with his 2016 presidential campaign, Donald Trump used "system-delegitimizing rhetoric and populist appeals to capitalize on existing fears and anxieties and create new ones, particularly focusing on anti-immigrant and racial appeals."[63]

Rhetoric aimed at delegitimizing the political system reached a fever pitch after the 2020 election, when Trump supporters operating under the banner "Stop the Steal" used Facebook, Twitter, and other platforms to try to undermine President Biden's victory. Republican members of Congress joined this anti-democratic campaign. On November 5, 2020, Rep. Andy Biggs, an Arizona Republican, retweeted Donald Trump Jr.'s declaration that "the best thing for America's future is for @realDonaldTrump to go to total war over this election to expose all of the fraud, cheating, dead/no longer in state voters, that has been going on for far too long."[64] Even after January 6, 147 Republican members of Congress voted to defy the results of a free and fair election.

More recently, in about a dozen states, Republican-controlled legislatures have invoked false accounts of election rigging in 2020 to justify new laws making it more difficult to vote. These measures, enacted in Florida, Georgia, Texas, and elsewhere, impose new limits on mail-in ballots, stiffen voter-identification requirements, reduce polling-site hours, cut back on the use of ballot drop boxes, and shift authority from career election officials to Republican-controlled legislatures. Although Republicans maintain that they're merely promoting "election integrity," the restrictions are widely expected to hinder participation by Democrats, especially Blacks and Latinos.[65]

Lessons From Abroad

Political volatility and social media make an explosive pair. Consider the 2021 military coup in Myanmar and the earlier ethnic cleansing of the country's Rohingya Muslim population; rising anti-Muslim sentiment in India and ethnic strife in Ethiopia; political harassment and misinformation in the Philippines and lethal hostility between Israelis and Palestinians. All of these combustible situations have been aggravated by social media. To avoid a worsening of our own divisions, Americans should study what has gone wrong abroad.

Two lessons stand out. First, in most of these countries, and others, the on-the-ground reality of how social media operates starkly contradicts Facebook's denials that its products contribute to political polarization. And second, social media can act as a political flamethrower, exacerbating existing tensions and, in some instances, creating new ones.

Access to social media has empowered dissidents, activists, and journalists around the globe, allowing them to call their governments to account. But since 2013, a number of governments have co-opted the platforms to serve their own ends. From Brazil to India and from Iran to Russia, repressive regimes have exploited social media to demonize political opponents.

Dividing the Philippines

Perhaps nowhere has social media's potential to divide people been more clearly demonstrated than in the Philippines. There, President Rodrigo Duterte has built a Facebook-powered propaganda machine that has intimidated critics and made a mockery of democratic aspirations.

It started with Duterte's successful 2016 presidential campaign, notable for his violent rhetoric on drug dealers. As his message spread on Facebook, Duterte's popularity grew. Local "influencers" hoping to capitalize on his new-found fame began distributing pro-Duterte propaganda to millions of Facebook followers. At one point in the campaign, nearly two-thirds of Facebook conversations in the Philippines were about him.

After he assumed the presidency, Duterte's massive online following morphed into an instrument of state power. Maria Ressa, founder of the Philippines' largest online news site, Rappler, watched with increasing concern as Duterte's government harnessed Facebook to silence critics and spread propaganda. In one particularly nasty episode, partisans of Duterte took down an opposition senator by spreading doctored pornographic images and lies about her personal finances.

After publishing an article critical of the government's online activities, Ressa herself became a target of Duterte's trolls. Fake accounts inundated her with hate mail, sending as many as 90 messages an hour. In 2019, she was arrested and later convicted on trumped-up charges of "cyber-libel," which carries a prison sentence of up to six years. In an interview, Ressa notes that before Duterte's dominance on Facebook, "All of us started pretty much in the center.

We [didn't] have polarization in news groups." Duterte changed that, one bot and falsehood at a time. Ressa remains in Manila, where she is free on bail, pending her appeal of her conviction.

Amplifying strife in the Middle East

The clash between Hamas and Israeli armed forces in spring 2021 provides another example of social media's capacity to exacerbate a volatile situation. In May, as Hamas rockets and Israeli Air Force bombs dropped, combatants on both sides turned to social media to whip up support and distort the facts. Misinformation in the form of videos, images, and text was shared thousands of times, including on Facebook, TikTok, Twitter, WhatsApp, and YouTube.

One post in Hebrew, reportedly shared via popular WhatsApp groups in Israel, raised the specter of an approaching Palestinian mob, stating, "Palestinians are coming, parents protect your children." A message in Arabic posted the same week to a large Palestinian WhatsApp group warned that Israeli soldiers were preparing to invade Gaza. Neither was true.

As it has done elsewhere, Facebook responded by surging emergency resources. The company established a "special operations center" in Israel staffed with native Arabic and Hebrew speakers tasked with monitoring for rule-violating content and restoring posts that had been improperly removed by automated systems. In a remarkable development, senior Facebook executives met virtually with Israeli and Palestinian officials to discuss how content moderation policies were affecting the conflict. Facebook and Twitter eventually acknowledged that they had wrongly blocked or restricted millions of mostly pro-Palestinian posts and accounts, in some cases because content-removal algorithms interpreted words like "martyr" and "resistance" as signaling calls to violence.

At its best, social media has democratized information across much of the world. As the Covid-19 pandemic raged in India in 2021, Facebook and its WhatsApp subsidiary helped doctors locate vital supplies and counter medical misinformation. But with these benefits come costs. Social media continues to allow dangerous falsehoods and inflammatory content to spread globally, including in India. With the insurrection at the U.S. Capitol—an event organized and promoted on social media platforms—many Americans experienced an unsettling example of what millions elsewhere in the world have lived with for years.

[1] https://www.lawfareblog.com/philippines-deserves-more-facebook; https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war

[2] https://www.bloomberg.com/news/features/2017-12-07/how-rodrigo-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook; https://www.rappler.com/nation/propaganda-war-weaponizing-internet

[3] https://www.nytimes.com/2021/05/14/technology/israel-palestine-misinformation-lies-social-media.html

[4] https://www.politico.eu/article/facebook-set-up-special-operation-center-for-content-related-to-israeli-palestinian-conflict/

[5] https://time.com/6050350/palestinian-content-facebook/; https://www.washingtonpost.com/technology/2021/05/28/facebook-palestinian-censorship/

Radicalization and violent extremism

The most ominous aspect of partisan hatred is that it can serve as a precursor of radicalization and violence, some of which is stoked and organized online. In the weeks leading up to January 6, Donald Trump riled up his supporters via Twitter and Facebook, repeatedly summoning them to Washington for a protest he promised "will be wild." In reaction to the ensuing mayhem, Twitter, Facebook, and YouTube took the extraordinary step of removing Trump's accounts from their platforms. Twitter did so permanently; Facebook, after some internal gyrations, imposed a two-year suspension. Both companies cited the danger that Trump would continue to incite violence. YouTube has said that it will reinstate Trump if the threat of civil unrest recedes.[66]

In an internal analysis, Facebook acknowledged shortcomings in its response to online signals that pro-Trump forces would try to disrupt congressional certification of the electoral votes on January 6. Content moderators focused too much on individual users, rather than coordinated networks seeking to undermine the election, according to the analysis, which *BuzzFeed News* obtained and published. The analysis noted that Facebook's policies stress the removal of "inauthentic" actors and content, such as the Russian operatives who disguised themselves as Americans and interfered with the 2016 presidential election. "What do we do when a movement is authentic, coordinated through grassroots or authentic means, but is inherently harmful and violates the spirit of our policy?" the Facebook report asked. "What do we do when that authentic movement espouses hate or delegitimizes free elections?" The document noted that an internal task force on "disaggregating harmful networks" is addressing these questions.[67]

> “In recognition of dangerous political fanaticism on social media, Facebook in mid-2021 began testing notifications to certain users that ask whether someone they know is leaning toward extremism. ‘Violent groups try to manipulate your anger and disappointment,’ one alert says, pointing users toward anti-extremism resources.”

Heightened affective polarization has fostered anti-government extremism in the U.S. among both far-left anarchists and right-wing militia groups. The looting and violence that accompanied some protests after George Floyd’s murder in 2020 signal the possibility of future unrest driven by the political left, according to Robert Pape, a political science professor at the University of Chicago who studies political violence. But at present, the threat comes primarily from the right, Pape says in an interview. In an “intelligence assessment” released publicly in May 2021, the Federal Bureau of Investigation and the Department of Homeland Security asserted that racially or ethnically motivated extremists who believe in the superiority of the white race are the primary sources of significant violence at present. The agencies further noted that domestic attacks are typically the work of “lone offenders, often radicalized online.”[68] In June 2021, the FBI and DHS issued a joint warning specifically about QAnon. As online predictions of Donald Trump’s reinstatement as president have failed to come true, QAnon adherents, who until then had seen themselves as “digital soldiers,” may initiate real-world violence, the agencies warned.[69]

An FBI investigation in early 2021 into a nascent plot to blow up Democrats’ California headquarters in Sacramento illustrates the potential danger. The probe led to charges against two men whose online communications indicated that they believed the 2020 election had been stolen from Trump and hoped that by committing violent acts they would incite a larger conflict. “I want to blow up a democrat building bad,” wrote one man, who allegedly had amassed an arsenal of 49 firearms, thousands of rounds of ammunition, and five pipe bombs.[70]

Even some Republicans now fear violence from those further to the right. In Clark County, Nevada, the local Republican Party has been shaken by an insurgent group reportedly associated with the Proud Boys, a militant pro-Trump organization. In May 2021, the county GOP canceled a meeting because of what it called the threat of physical danger.[71] Members of the Proud Boys, along with other extremist groups, such as the Oath Keepers and Three Percenters, face felony charges related to the Capitol insurrection.[72] Around the country, state and local election officials, Republicans as well as Democrats, have received death threats from Americans angry about Trump’s defeat in November 2020. In Georgia, for example, these threats have referred to “hanging, firing squads, torture, and bomb blasts.”[73]

In mid-2021, Facebook began testing notifications to certain users that ask whether someone they know is leaning toward extremism. One of the alerts says: “Violent groups try to manipulate your anger and disappointment. You can take action now to protect yourself and others.” The notices direct users to a variety of resources, including Life After Hate, a group that helps people leave violent far-right movements.[74]

‘Obligation to protect democracy’

Some analysts argue that reducing overall political polarization deserves less emphasis than trying to stamp out white supremacy. According to this view, people of color suffer disproportionately from asymmetric democratic erosion and extremism. Daniel Kreiss and Shannon McGregor, both researchers at the Center for Information, Technology, and Public Life at the University of North Carolina, argued in an April 2021 commentary in *Wired* that “it is entirely understandable [that] many people of color and people of all races committed to equality would have negative feelings about the Republican partisans on the ‘other side.’”[75] In an interview, McGregor adds: “The problem is not that [American society and its politics are] polarized. The problem is that one pole of the polarization is behaving anti-democratically and illiberally."

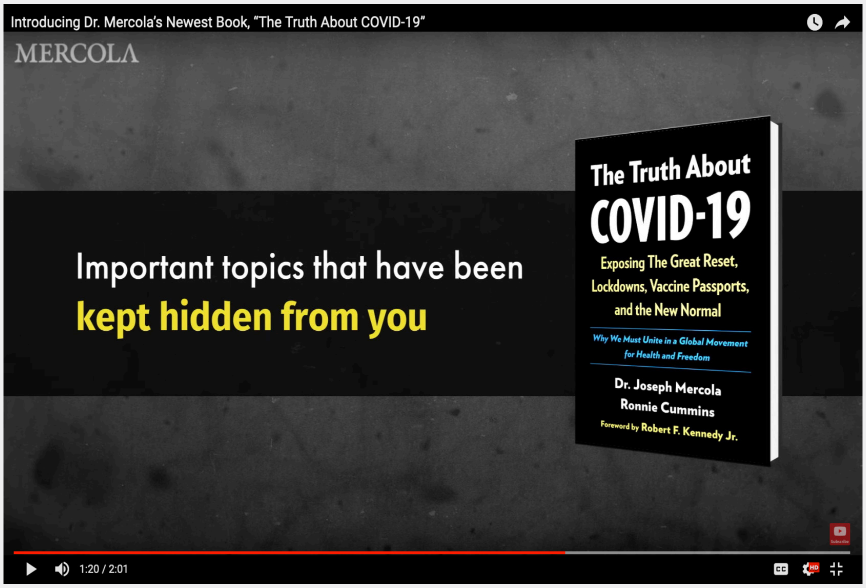
There is broad expert agreement that racism and fear of loss of socio-economic status animate the anti-democratic extremism of many conservative white Americans. Jennifer McCoy, the Georgia State University political scientist, argues that the most extreme polarization in the U.S. can be traced to a "group of white Christian males who view their status in society as decreasing."[76] In a June 2021 paper that analyzes results of four surveys of thousands of respondents between 2011 and 2018, political scientists Lilliana Mason, Julie Wronski, and John Kane identify "a wellspring of animus against marginalized groups in the United States that can be harnessed for political gain." Support for Donald Trump, they add, is "uniquely tied to animus toward minority groups."[77] Robert Pape's 2021 surveys of self-identified U.S. conservatives identify fear of a "great replacement"—meaning whites being eclipsed by non-whites—as a key motivation for support of the January 6 insurrection. Pape found that 50% of Republicans "believe non-whites will have more rights than them in the future, as compared to 16% of non-Republicans who hold this belief."[78]

In this context, social media companies have "an ethical obligation to protect democracy and human rights," says North Carolina's Kreiss. "The public should expect that from them. I think lawmakers and journalists and other stakeholders should expect that of them. And I think that every decision that they make, when it comes down to how they design, interpret, and enforce their content policies, they should come with that in mind: Are we furthering things like the peaceful transfer of power? Are we furthering things like the legitimacy of elections? Are we ensuring that the speech on our platform is not dehumanizing certain groups of people or individuals?" This report's goals—clarifying the relationship between social media and polarization, identifying the consequences of extreme divisiveness, and offering recommendations for reducing polarization—are not at odds with these aims.

Like Kreiss and McGregor, Jonathan Stray, a researcher at the University of California, Berkeley Center for Human-Compatible Artificial Intelligence, emphasizes that the pursuit of social justice can naturally increase polarization. Stray suggests that social media platforms need to experiment with affirmative "depolarization" to avoid violence and address deep-seated racism. Alluding to ideas drawn from the field of "peace building," he sees "the goal of depolarization as conflict transformation: not eliminating or resolving conflict but making conflict better in some way, e.g. less prone to violence and more likely to lead to justice." He suggests that it's possible to address polarization and its asymmetric consequences across the full range of American political issues without losing sight of social justice as a top priority.

Divisive content

More than 400,000 people subscribe to the YouTube channel of Dr. Joseph Mercola, whom researchers have labeled a leading source of coronavirus misinformation. Here, a video advertising his 2021 book, "The Truth About Covid-19," warns that "the technocratic overlords" are using the pandemic to "eliminate your privacy and personal liberties." Mercola, who also has large followings on Facebook and Twitter, has accused his critics of trying to censor his efforts to publicize alternative health products, which he sells online.


Introducing Dr. Mercola's Newest Book, "The Truth About COVID-19"
MERCOLA
Important topics that have been
kept hidden from you
The Truth About
COVID-19
Exposing The Great Reset, Lockdowns, Vaccine Passports, and the New Normal
Why We Must Unite in a Global Movement for Health and Freedom
Dr. Joseph Mercola
Ronnie Cummins
Foreword by Robert F. Kennedy Jr.
1:20 / 2:01

Part 3: Conclusion and Recommendations

In the U.S., the social media industry operates in a democracy suffering the consequences of extreme political polarization. Having exacerbated this divisiveness, the major social media companies have a responsibility, therefore, to make changes that will ease partisan hatred and help begin to repair the damage they have done. This challenge is now so large and complicated that it will require the intervention of the government, as well.

> “The very polarization to which both social media platforms and political leaders have contributed will make it difficult to achieve progress in Washington. But that’s not an excuse for inaction.”

The threat to American democracy is real. In its December 2020 “trust barometer” survey, Edelman found that a majority of both Republicans and Democrats—57% of total respondents—agree with the statement, “The degree of political and ideological polarization in this country has gotten so extreme that I believe the U.S. is in the midst of a cold civil war.”[79] In this environment, the risk of further political violence is great. According to recent polling by the Public Religion Research Institute, 15% percent of Americans—about 50 million people—agree that “because things have gotten so far off track, true American patriots may have to resort to violence in order to save our country.” Republicans, at 28%, are four times more likely than Democrats to agree.[80]

Alongside the peril of political violence is the continuing threat to democratic participation. In June 2021, more than 100 academic experts on democracy brought together by the liberal think tank New America issued an alarm about Republican-led state legislatures that are pursuing “radical changes to core electoral procedures in response to unproven and intentionally destructive allegations of a stolen election.”[81] Just days later, a conservative six-member majority of the Supreme Court made it significantly more difficult to win lawsuits alleging violations of the federal Voting Rights Act of 1965. “This is a do-or-die moment for American democracy,” Hakeem Jefferson, a political scientist at Stanford, says in an interview. “Perhaps I sound alarmist in our conversation, because I think it is a moment in which we should all be alarmed.”

Social media companies cannot rescue the United States from itself. But these companies can, and must, reform their practices when they cause harm to democracy. In light of the industry’s failure to engage in sufficiently vigorous self-regulation, however, it is now time for the government to step in, as well. The very polarization to which both social media platforms and political leaders have contributed will make it difficult to achieve progress in Washington. But that’s not an excuse for inaction.

Recommendations to the federal government:

President Biden:

1 Prioritize a broad government response to the heightening of partisan hatred by social media.

For years, Washington politicians have debated and castigated social media without coherently addressing its role in fostering political polarization. This needs to change. For the moment, though, Congress seems unable to overcome the very sort of dysfunction that is one of the consequences of extreme political division and distrust. Responsibility falls to President Biden to make these issues a national priority—and he needs to do so in a serious, deliberate way. Unfortunately, his remark in July 2021 that Facebook is "killing people" by spreading misinformation about Covid-19 vaccinations oversimplified a complicated problem, and Biden's subsequent attempt to walk back the comment only added to the confusion. In contrast, Surgeon General Vivek Murthy earlier had issued a formal health advisory urging social media companies to clamp down on vaccine misinformation. Murthy's well-reasoned warning was a good first step; now the president himself needs to lend his authority to the cause in a clear and compelling manner.[82]

Biden has options: In one or more speeches, by means of a bipartisan blue-ribbon commission, or via some other high-visibility vehicle, he should tell both lawmakers and the public that to avoid the politicization of public health crises and future versions of the Capitol insurrection, we must confront online polarization and its malign effects. By demonstrating leadership in this fashion, Biden can begin to break the logjam in Congress and open a path for achieving other goals outlined here.

Congress:

2 Investigate the role of social media in the January 6 insurrection.

In establishing a select committee to probe the causes of the invasion of the Capitol, House Speaker Nancy Pelosi included important language on investigating how technology was used to incite the insurrection. The committee has a crucial opportunity to shed light on the consequences of partisan hatred and how the interaction between social media, hyper-partisan news media, political leaders, and protesters motivated the violence on January 6. Panel members must make this a central line of inquiry and use their subpoena power to pursue it. Facebook's quasi-independent Oversight Board has urged the company itself to investigate these matters. But Facebook management declined, saying that Congress should assume the responsibility. Now, lawmakers must do so.[83]

3 Mandate more disclosure about the inner workings of social media platforms, so outside researchers can analyze the data.

It's difficult to propose specific remedies for many of the problems associated with social media because the companies refuse to disclose how their platforms work. "We do not know even what we do not know concerning a host of pathologies attributed to social media and digital communication technologies," Nathaniel Persily, a law professor at Stanford, wrote recently.[84]

Congress should address this data deficit by requiring more transparency, but with sensible legal protection for both the companies and qualified researchers. Persily has proposed legislation that would do three things: First, it would compel the largest platforms, namely Facebook and Google/YouTube, to share data on how algorithms rank, recommend, and remove content. Second, it would protect the platforms from civil and criminal liability when they share this information with vetted academics under prescribed circumstances. And third, it would legally immunize researchers when they use the data. This approach would facilitate more in-depth research, which, in turn, could lead to better-informed public policies.

Rebekah Tromble, director of the Institute for Data, Democracy & Politics and an associate professor at George Washington University, is leading a parallel discussion on data access via a regulatory working group in Europe. Tromble hopes the E.U. will create a voluntary corporate code of conduct that would encourage disclosure. It could be incorporated into regulatory regimes in Europe and, eventually, the U.S. The Federal Trade Commission, Tromble suggests, might be the appropriate U.S. agency to oversee new transparency rules in this country.

4 Empower the Federal Trade Commission to draft and enforce an industry code of conduct.

The FTC's oversight of social media needs to go much further than data disclosure. We urge Congress to pass legislation authorizing the agency to collaborate with social media companies and other stakeholders to create standards for industry conduct that would be enforceable by the government.

The standards would define the duties of social media companies when addressing hateful, extremist, or threatening content. In addition to data transparency, the standards could set benchmarks for the amount of various categories of harmful content that remains on platforms even after automated and human moderation. If the benchmarks are exceeded, fines could be imposed. The standards could also require minimum protections of user privacy. Working with colleagues at the Harvard Kennedy School, we proposed such standards earlier this year in recommendations made to the Biden Administration.[85]

While it would make sense for industry representatives to bring their technical expertise to the task of drafting the standards, the government would have the ultimate authority to approve and enforce them. Congress could require social media companies to incorporate the new rules into their terms-of-service agreements with users. Then, if the companies fail to observe the standards, the FTC could initiate enforcement action under its existing authority to police "unfair or deceptive" commercial practices.

Democratic Representatives Jan Schakowsky of Illinois and Kathy Castor of Florida have introduced a bill that points generally in the direction we're recommending. Their Online Consumer Protection Act would require social media companies to incorporate into their terms of service how they handle mis- and disinformation related to public health and elections, among other topics, and to establish a program ensuring compliance with consumer protection laws. The FTC, state attorneys general, and individual plaintiffs would have the ability to go to court to enforce these requirements.[86]

5 Encourage exploration of alternatives to current social media business models.

While it grapples with social media as it now exists, Congress should provide research funding that encourages technologists and entrepreneurs who are imagining a radically different, pro-democratic digital future. Given the dominant market positions of the incumbent companies and their penchant for acquiring or overwhelming smaller competitors, public support is necessary to nurture alternatives.

One worthy idea is the development of "public service digital media," as scholars such as Ethan Zuckerman at the University of Massachusetts, Amherst have proposed. The goal is to build sites that operate primarily to promote civic values rather than profits: a Public Broadcasting System of the internet. "Instead of optimizing for raw engagement, networks like these would measure success in terms of new connections, sustained discussions, or changed opinions," Zuckerman has suggested.[87] Eli Pariser, who leads New Public, which promotes the design of digital public spaces, points to local nonprofit experiments, such as the Vermont-based Front Porch Project. A heavily moderated listserv that fosters civil discussion, Front Porch boasts participation by two-thirds of Vermont households.[88]

Members of a Stanford working group advocate a dramatic overhaul of existing social media platforms. They propose separating the basic social networks that billions of people have joined from the algorithmic functions of ranking and moderating content. According to this new model, Facebook or YouTube members would be able to choose from a marketplace of smaller firms offering a variety of approaches to determining who sees what content. The aim is to reduce the influence of any one social media company on public discourse and democracy.[89] At least one Silicon Valley tycoon is intrigued by such thinking. Jack Dorsey, the founder and CEO of Twitter, has launched an initiative called Bluesky to explore replacing centralized platforms with interoperable components.[90]

As appealing as all of this may sound, it's far from clear why companies like Facebook or Google, which are much larger and more lucrative than Twitter, would embrace new approaches that, by definition, would reduce the profitability of their current advertising-driven business models. That's all the more reason why government should provide incentives to those who are pursuing potentially constructive alternatives to today's polarization-inducing social media industry.

Recommendations to the platforms:

6 Adjust algorithms to depolarize platforms more systematically.

Social media companies that currently accelerate polarization should deploy their engineering prowess to do the opposite. In Part 1, we described emergency episodes related to the November 2020 election and the April 2021 Derek Chauvin trial in which Facebook temporarily modified its algorithms. These measures sought to decrease the reach of politically polarizing and extremist content, while favoring more authoritative information about controversial issues. We recommend application of such measures in a more systematic way, not just in anticipation of potential crises.

After January 6, Mark Zuckerberg announced that Facebook would make permanent a provisional decision to stop recommending to users that they join politically oriented Groups. The platform's recommendation algorithm had been steering some people toward pockets of hyper-partisan antagonism, including Groups promoting QAnon and Stop the Steal. Zuckerberg called the move "a continuation of work we've been doing for a while to turn down the temperature and discourage divisive conversations and communities."[91] That's fine, as far as it goes. But in addition to curbing recommendations, the major platforms need to filter out the most harmful, polarizing content on an ongoing basis. In anticipation of this shift, the platforms should devote significant additional resources to refining their moderation algorithms in order to minimize removal of material that, upon closer inspection, isn't problematic. Mistakes inevitably will occur, but they must be kept to a minimum and, when identified, promptly corrected.

Jonathan Stray, the Berkeley researcher, theorizes that platforms should be able to go even further. He suggests that they could create metrics to track surges in affective polarization and then respond with algorithmic adjustments designed to elevate the terms of online conflict and thus ease partisan hatred. Changes of this sort worth exploring include recommending more civil, constructive arguments that users could consider. Platform product designers could even rely on polarization metrics to anticipate whether new features may exacerbate partisan hatred.[92]

7 Make depolarizing adjustments more transparent.

If they follow our recommendation to step up filtering of polarizing content, it's imperative that the platforms be much more open about what they're doing, how they're doing it, and what content might potentially get blocked in the process. Transparency is the only way to counter suspicions that such measures are designed to manipulate politics or otherwise exert illegitimate influence.

With respect to the example involving Facebook Groups, mentioned in the previous recommendation, the company needs to reveal more about how certain Groups have become havens for extremism and what it is doing to address the problem. As social media companies continue to improve the design of their platforms to diminish the amplification of partisan hatred, they will need to do so in a way that allows users, and society at large, to assess their effectiveness and hold them accountable.

8 Double the number of human content moderators and bring them in-house.

Facebook and other social media companies face an increasingly daunting challenge in policing the billions of written statements, still images, and videos posted on their platforms every day. Continuing to refine content moderation algorithms is one necessary response, but it's not enough. As impressive as it can be, artificial intelligence struggles to assess context: Is a call to violence posted to inspire insurrection, or is it offered to condemn extremism? This is where human moderation remains critical. In recent years, the companies have expanded the number of people moderating content, but these ranks need to grow much more. As we noted in a report last year, if Facebook doubled its moderation workforce to 30,000, front-line reviewers would have more time to consider difficult content decisions. A larger moderator corps would also allow supervisors to rotate assignments more frequently so that reviewers exposed to the most disturbing content could switch to less brutal material.[93]

And that raises a related point: At present, the vast majority of content moderation is farmed out to third-party contractors in places like the Philippines, Ireland, and India. The companies should bring content moderation in-house, treating it as a core business function and assigning the task only to full-fledged employees. It's worth noting that the video platform TikTok already hires its moderators as in-house employees. The added expense of this approach simply reflects the true cost of doing business responsibly as a global social media platform.

9 Strengthen engagement with civil society groups that can help identify sources of harmful content.

The problem of online disinformation is so enormous that in recent years, civil society has cobbled together its own patchwork response. In the U.S., a variety of partnerships emerged in 2020 to contend with disinformation about the presidential election, Covid-19, and racial harassment. Platforms ought to expand their collaboration with groups such as the Election Integrity Partnership, which brought together university, civil society, and private sector researchers to identify false claims about voting. Other examples include the Disinformation Defense League, a partnership of more than 200 civil society organizations working to combat hate speech and disinformation, and the Virality Project, which has targeted Covid-19 disinformation.[94]

Social media companies should help these initiatives by introducing new ways for them to share information with the platforms and one another. But the companies must carefully assess the agendas and relative capabilities of those offering assistance. Many Palestinians, for example, have complained that they have experienced an unjustified degree of censorship on social media because the Israeli government has a proficient cyber unit that flags large quantities of allegedly hateful and violent Palestinian content. The Palestinians lack comparable capacity.[95] This imbalance doesn't mean that platforms should disregard Israeli reports; it means that these alerts, as well as those from Palestinian sources, need to be evaluated carefully.

10 Diminish rewards for virality and performative politics.

Social media users' eagerness to see their posts go viral leads to the spread of extreme, divisive content and what has come to be called "performative politics." A number of researchers, including Jaime Settle of William & Mary, Jonathan Haidt of NYU, and José Marichal of California Lutheran University, argue that social media companies ought to remove or downplay platform features that may contribute to polarizing online performances. This is a promising idea. Facebook and Twitter could obscure like and share counts, Haidt writes, "so that individual pieces of content can be evaluated on their own merit, and so that social media users are not subject to continual public popularity contests."[96]

Damon Centola, a sociologist at the University of Pennsylvania, argues that it's possible to design platforms so that instead of contributing to polarization, online interaction "actually creates further agreement and better understanding." In an interview, he describes how tweaking certain design features, such as the sort of information that readers use to infer the identity of posters, can lead to more consensus. His experiments have found, for example, that providing fewer specifics about posters' identity tends to diminish sensationalistic online behavior.[97]

Facebook has taken steps in this direction. It announced in May 2021 that it would give users of the Facebook and Instagram platforms the option of hiding their like counts, a move designed to "depressurize people's experience."[98] Diminishing the competition over who can generate posts that go viral might make social media less fun for some users. But that seems like a small price to pay for taking away tools that bad actors rely on to heighten toxicity and polarization.

Endnotes

[1] https://www.buzzfeednews.com/article/ryanmac/facebook-execs-polarization-playbook

[2] https://www.rev.com/blog/transcripts/mark-zuckerberg-opening-statement-transcript-house-hearing-on-misinformation

[3] https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2

[4] https://www.rev.com/blog/transcripts/senate-hearing-on-social-media-algorithms-full-transcript-april-27. The social media platforms present themselves in an entirely different light. At the Senate hearing in April 2021, Alexandra Veitch, director of government affairs and public policy for the Americas and emerging markets for Google's YouTube platform, said: "YouTube's business relies on the trust of our users, our creators and our advertisers. That's why responsibility is our number one priority."

[5] https://www.nytimes.com/2021/05/05/technology/facebook-trump-nick-clegg.html?action=click&module=Top%20Stories&pgtype=Homepage

[6] https://science.sciencemag.org/content/370/6516/533; https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(21)00196-0

[7] https://about.fb.com/news/2020/05/investments-to-fight-polarization/

[8] https://www.aapss.org/volumes/polarizing-polities-a-global-threat-to-democracy/

[9] https://science.sciencemag.org/content/370/6516/533

[10] https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763/

[11] https://apnews.com/article/media-immigration-donald-trump-minnesota-ap-top-news-e43cf06befa24408b5a500626f2550d9; https://abcnews.go.com/US/trumps-language-mexican-immigrants-scrutiny-wake-el-paso/story?id=64768566; https://www.brookings.edu/techstream/how-hate-and-misinformation-go-viral-a-case-study-of-a-trump-retweet/aso/story?id=64768566; https://www.washingtonpost.com/politics/trump-lashes-out-at-black-lives-matter-accuses-one-member-of-treason/2020/06/25/45667ec8-b70f-11ea-a510-55bf26485c93_story.html

[12] Some scholars have questioned the widely held assumption that extreme affective polarization tends to undermine democracy. See, e.g., https://osf.io/9btsq/ ("affective polarization's consequences should be generally confined to interpersonal domains,with more circumscribed political implications").

[13] https://bhr.stern.nyu.edu/

[14] https://www.aeaweb.org/articles?id=10.1257/aer.20190658. A study published in June 2021 found that when Facebook users in Bosnia and Herzegovina stayed off the platform for a week, their regard for other ethnic groups fell. This finding suggests that in an ethnically riven country like Bosnia and Herzegovina, social media use may reduce polarization, perhaps because information available offline is more divisive than what is available online: https://drive.google.com/file/d/1Vqgo2HYbBYoExnf_0AiH2560BY5zbXMf/view

[15] https://www.pnas.org/content/114/40/10612

[16] https://journals.sagepub.com/doi/full/10.1177/0002716218811309

[17] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3522318

[18] https://www.vox.com/2020/1/24/21076232/polarization-america-international-party-political; https://simonandschusterpublishing.com/why-were-polarized/

[19] https://science.sciencemag.org/content/370/6516/533

[20] https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx; https://www.statista.com/statistics/266206/googles-annual-global-revenue/

[21] https://www.cambridge.org/core/books/frenemies/00D051D46BC4CDB2D322EE6A1CEA6791

[22] https://www.washingtonpost.com/nation/2020/05/13/protest-violence-coronavirus/; https://www.politico.com/news/2021/06/05/partisan-divide-vaccinations-491947

[23] https://about.fb.com/news/2018/01/sunstein-democracy/. See also Sunstein's 2017 book *#Republic: Divided Democracy in the Age of Social Media*: https://press.princeton.edu/books/hardcover/9780691175515/republic

[24] https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20191777

[25] https://science.sciencemag.org/content/348/6239/1130

[26] https://www.pnas.org/content/115/37/9216

[27] https://press.princeton.edu/books/hardcover/9780691203423/breaking-the-social-media-prism

[28] https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

[29] https://d1gi.medium.com/untrue-tube-monetizing-misery-and-disinformation-388c4786cc3d

[30] https://www.pnas.org/content/118/32/e2101967118.short?rss=1

[31] https://blog.youtube/inside-youtube/building-greater-transparency-and-accountability/

[32] https://www.youtube.com/intl/en-GB/about/press/

[33] https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499?mod=hp_lead_pos5

[34] https://about.fb.com/news/2020/05/investments-to-fight-polarization/

[35] https://twitter.com/ylecun/status/1372673301902934017?lang=en

[36] https://about.fb.com/news/2021/03/changes-to-keep-facebook-groups-safe/; https://www.nasdaq.com/articles/facebook-starts-to-remove-recommendations-for-political-and-social-groups-globally-2021-03

[37] https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

[38] https://thehill.com/changing-america/respect/equality/527384-facebook-employees-propose-post-election-changes-to-reduce; https://www.harpercollins.com/products/an-ugly-truth-sheera-frenkelcecilia-kang?variant=32999376551970

[39] https://www.harpercollins.com/products/an-ugly-truth-sheera-frenkelcecilia-kang?variant=32999376551970

[40] https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election/; https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/; https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html

[41] https://about.fb.com/news/2021/04/preparing-for-a-verdict-in-the-trial-of-derek-chauvin/

[42] https://www.theatlantic.com/ideas/archive/2021/04/facebook-should-dial-down-toxicity-much-more-often/618653/

[43] https://www.rev.com/blog/transcripts/senate-hearing-on-social-media-algorithms-full-transcript-april-27

[44] https://twitter.com/ReliableSources/status/1416820494045679618

[45] https://www.basicbooks.com/titles/thomas-e-mann/its-even-worse-than-it-looks/9780465096206/

[46] https://docs.cdn.yougov.com/pxuc7wjg52/econTabReport.pdf

[47] https://www.cbsnews.com/news/january-6-opinion-poll/

48 https://www.wsj.com/articles/why-are-americans-so-distrustful-of-each-other-11608217988

49 https://www.pewresearch.org/politics/2019/10/10/the-partisan-landscape-and-views-of-the-parties/

50 https://www.edelman.com/trust/2021-trust-barometer

51 https://www.buzzfeednews.com/article/drumoorhouse/qanon-mass-collective-delusion-buzzfeed-news-copy-desk

52 https://www.npr.org/2021/01/31/962104747/unwelcome-on-facebook-twitter-qanon-followers-flock-to-fringe-sites. In an academic literature review for his 2020 book, *Trust in a Polarized Age*, Kevin Vallier, an associate professor of philosophy at Bowling Green State University, found little evidence to date of a direct connection between social media and distrust. This would make for a fruitful area for future research. https://global.oup.com/academic/product/trust-in-a-polarized-age-9780190887223?cc=us&lang=en&.

53 https://www.nytimes.com/2021/05/29/us/gun-purchases-ownership-pandemic.html

54 https://www.bloomberg.com/news/features/2021-07-07/where-is-delta-spreading-u-s-midwest-rockies-as-trump-country-rejects-vaccine

55 https://www.nytimes.com/2021/05/07/world/asia/misinformation-disinformation-fake-news.html

56 https://advances.sciencemag.org/content/7/23/eabf1234

57 https://www.brookings.edu/techstream/how-partisan-polarization-drives-the-spread-of-fake-news/

58 https://cshe.berkeley.edu/polarization-trust-american-institutions-1970; https://news.gallup.com/poll/352397/democratic-republican-confidence-science-diverges.aspx

59 https://academic.oup.com/joc/article/69/2/168/5425470

60 https://www.amazon.com/Polarization-What-Everyone-Needs-Know%C2%AE/dp/0190867779

61 https://www.pnas.org/content/118/26/e2024292118

62 https://www.aft.org/ae/fall2020/levitsky_ziblatt

63 https://journals.sagepub.com/doi/full/10.1177/0002716218818782

64 https://techpolicy.press/rep-zoe-lofgren-publishes-analysis-of-social-media-posts-from-102-republicans-who-voted-to-overturn-the-2020-election/

65 https://www.brennancenter.org/our-work/research-reports/voting-laws-roundup-may-2021

66 https://blog.twitter.com/en_us/topics/company/2020/suspension; https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/; https://www.politico.com/news/2021/03/04/youtube-trump-reinstate-risk-of-violence-473716

67 https://www.buzzfeednews.com/article/ryanmac/full-facebook-stop-the-steal-internal-report

68 https://www.fbi.gov/investigate/terrorism

69 https://www.cnn.com/2021/06/14/politics/fbi-qanon-warning-to-lawmakers/index.html

70 https://www.cnn.com/2021/07/16/politics/democratic-headquarters-sacramento-plot/index.html

71 https://www.thedailybeast.com/clark-county-nevada-gop-cancels-meeting-amid-fear-of-proud-boy-insurgency?via=twitter_page

72 https://www.usatoday.com/story/news/2021/03/24/capitol-attack-oath-keepers-proud-boys-three-percenters-coordinated/6980128002/

73 https://www.reuters.com/investigates/special-report/usa-trump-georgia-threats/

74 https://www.cnn.com/2021/07/01/tech/facebook-extremist-notification/index.html

75 https://www.wired.com/story/polarization-isnt-americas-biggest-problem-or-facebooks/

76 https://journals.sagepub.com/doi/full/10.1177/0002716218818782; https://www.youtube.com/watch?v=6yk6bnwFaQk

77 https://www.cambridge.org/core/journals/american-political-science-review/article/activating-animus-the-uniquely-social-roots-of-trump-support/D96C71C353D065F62A3F19B504FA7577

78 https://www.uchicago.edu/research/center/the_chicago_project_on_security_and_threats/

79 https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer.pdf

80 https://www.prri.org/research/qanon-conspiracy-american-politics-report/

81 https://www.newamerica.org/political-reform/statements/statement-of-concern

82 https://apnews.com/article/joe-biden-business-health-media-social-media-73ca875f1d1c04bc69108607d8499e3c

83 https://techpolicy.press/facebook-says-elected-officials-should-investigate-its-role-in-january-6/

84 https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/cpc-open_windows_np_v3.pdf

85 https://bit.ly/3hhKTQl

86 https://schakowsky.house.gov/media/press-releases/schakowsky-castor-introduce-online-consumer-protection-act

87 https://www.cjr.org/special_report/building-honest-internet-public-interest.php

88 https://www.politico.com/news/agenda/2021/01/05/to-thrive-our-democracy-needs-digital-public-infrastructure-455061

89 https://www.foreignaffairs.com/articles/united-states/2020-11-24/fukuyama-how-save-democracy-technology

90 https://www.theverge.com/2021/1/21/22242718/twitter-bluesky-decentralized-social-media-team-project-update

91 https://investor.fb.com/investor-events/default.aspx

92 https://arxiv.org/abs/2107.04953

93 https://bit.ly/3h56BYZ

94 https://www.eipartnership.net/; https://www.protocol.com/tag/disinformation-defense-league; https://www.viralityproject.org/

95 https://techpolicy.press/social-media-conflict-and-censorship-in-palestine/

96 https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763/

97 https://penntoday.upenn.edu/news/climate-change-political-polarization-disappears-social-networks

98 https://about.instagram.com/blog/announcements/giving-people-more-control

Appendix

People Interviewed for This Report

Christopher Bail
Duke University

Michael Beckerman
TikTok

Levi Boxell
Stanford University

Damon Centola
University of Pennsylvania

Tim Colbourne
Facebook

Renée DiResta
Stanford University

Evelyn Douek
Harvard University

Hany Farid
University of California, Berkeley

Matthew Gentzkow
Stanford University

Dipayan Ghosh
Harvard University
Formerly Facebook

Zachary Graves
Lincoln Network

Andrew Guess
Princeton University

Eric Han
TikTok

Katie Harbath
Anchor Change
Formerly Facebook

Jonathan Haidt
New York University

Jess Hemerly
YouTube

Jason Hirsch
Facebook

Hakeem Jefferson
Stanford University

Daphne Keller
Stanford University
Formerly Google

Karen Kornbluh
German Marshall Fund of the United States

Daniel Kreiss
University of North Carolina

Yann LeCun
New York University
Facebook

José Marichal
California Lutheran University

Shannon McGregor
University of North Carolina

Filippo Menczer
Indiana University

Mor Naaman
Cornell University

Mutale Nkonde
AI for the People

Dawn Nunziato
George Washington University

Robert Pape
University of Chicago

Matt Perault
Duke University
Formerly Facebook

Nathaniel Persily
Stanford University

Nick Pickles
Twitter

Fadi Quran
Avaaz

Maria Ressa
Rappler

Chris Riley
R Street Institute

Geoff Samek
YouTube

John Samples
Cato Institute
Facebook Oversight Board

Marietje Schaake
Stanford University

Jaime Settle
College of William & Mary

Kate Starbird
University of Washington

Jonathan Stray
University of California, Berkeley

Talia Stroud
University of Texas, Austin

Rebekah Tromble
George Washington University

Joshua Tucker
New York University

Siva Vaidhyanathan
University of Virginia

Clement Wolf
Google

NYU | STERN

Center for Business
and Human Rights