ANNALS OF INQUIRY

HOW HARMFUL IS SOCIAL MEDIA?

There's a general sense that it's bad for society—which may be right. But studies offer surprisingly few easy answers.

By Gideon Lewis-Kraus

June 3, 2022

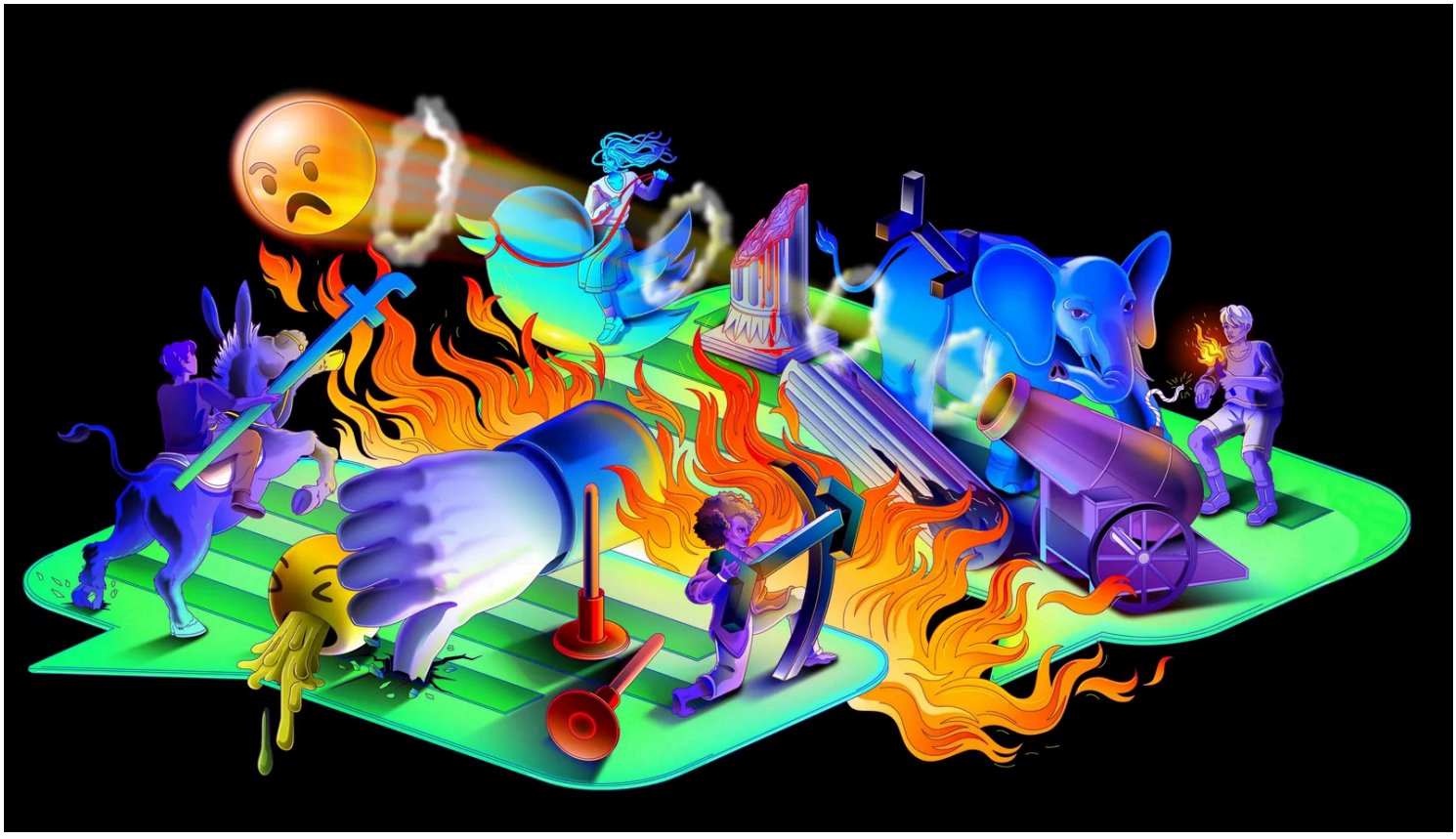


Illustration by Ari Liloan

Save this story

In April, the social psychologist Jonathan Haidt published an essay in *The Atlantic* in which he sought to explain, as the piece's title had it, "Why the Past 10 Years of American Life Have Been Uniquely Stupid." Anyone familiar with Haidt's work in the past half decade could have anticipated his answer: social media. Although Haidt concedes that political polarization and factional enmity long predate the rise of the platforms, and that there are plenty of other factors involved, he believes that the tools of virality—Facebook's Like and Share buttons, Twitter's Retweet function—have algorithmically and irrevocably corroded public life. He has determined that a great historical discontinuity can be dated with some precision to the period between 2010 and 2014, when these features became widely available on phones.

"What changed in the 2010s?" Haidt asks, reminding his audience that a former Twitter developer had once compared the Retweet button to the provision of a four-year-old with a loaded weapon. "A mean tweet doesn't kill anyone; it is an attempt to shame or punish someone publicly while broadcasting one's own virtue, brilliance, or tribal loyalties. It's more a dart than a bullet, causing pain but no fatalities. Even so, from 2009 to 2012, Facebook and Twitter passed out roughly a billion dart guns globally. We've been shooting one another ever since." While the right has thrived on conspiracy-mongering and misinformation, the left has turned punitive: "When everyone was issued a dart gun in the early 2010s, many left-leaning institutions began shooting themselves in the brain. And, unfortunately, those were the brains that inform, instruct, and entertain most of the country." Haidt's prevailing metaphor of thoroughgoing fragmentation is the story of the Tower of Babel: the rise of social media has "unwittingly dissolved the mortar of trust, belief in institutions, and shared stories that had held a large and diverse secular democracy together."

These are, needless to say, common concerns. Chief among Haidt's worries is that use of social media has left us particularly vulnerable to confirmation bias, or the propensity to fix upon evidence that shores up our prior beliefs. Haidt acknowledges that the extant literature on social media's effects is large and

complex, and that there is something in it for everyone. On January 6, 2021, he was on the phone with Chris Bail, a sociologist at Duke and the author of the recent book "Breaking the Social Media Prism," when Bail urged him to turn on the television. Two weeks later, Haidt wrote to Bail, expressing his frustration at the way Facebook officials consistently cited the same handful of studies in their defense. He suggested that the two of them collaborate on a comprehensive literature review that they could share, as a Google Doc, with other researchers. (Haidt had experimented with such a model before.) Bail was cautious. He told me, "What I said to him was, 'Well, you know, I'm not sure the research is going to bear out your version of the story,' and he said, 'Why don't we see?'"

Bail emphasized that he is not a "platform-basher." He added, "In my book, my main take is, Yes, the platforms play a role, but we are greatly exaggerating what it's possible for them to do—how much they could change things no matter who's at the helm at these companies—and we're profoundly underestimating the human element, the motivation of users." He found Haidt's idea of a Google Doc appealing, in the way that it would produce a kind of living document that existed "somewhere between scholarship and public writing." Haidt was eager for a forum to test his ideas. "I decided that if I was going to be writing about this—what changed in the universe, around 2014, when things got weird on campus and elsewhere—once again, I'd better be confident I'm right," he said. "I can't just go off my feelings and my readings of the biased literature. We all suffer from confirmation bias, and the only cure is other people who don't share your own."

Looking for new books, TV, and music? Get the Goings On newsletter for picks from our writers and critics.

gideonlk@gmail.com

SIGN UP

By signing up, you agree to our user agreement (including class action waiver and arbitration provisions), and acknowledge our privacy policy.

Haidt and Bail, along with a research assistant, populated the document over the course of several weeks last year, and in November they invited about two dozen scholars to contribute. Haidt told me, of the difficulties of social-scientific methodology, "When you first approach a question, you don't even know what it is. 'Is social media destroying democracy, yes or no?' That's not a good question. You can't answer that question. So what *can* you ask and answer?" As the document took on a life of its own, tractable rubrics emerged—Does social media make people angrier or more affectively polarized? Does it create political echo chambers? Does it increase the probability of violence? Does it enable foreign governments to increase political dysfunction in the United States and other democracies? Haidt continued, "It's only after you break it up into lots of answerable questions that you see where the complexity lies."

Haidt came away with the sense, on balance, that social media was in fact pretty bad. He was disappointed, but not surprised, that Facebook's response to his article relied on the same three studies they've been reciting for years. "This is something you see with breakfast cereals," he said, noting that a cereal company "might say, 'Did you know we have twenty-five per cent more riboflavin than the leading brand?' They'll point to features where the evidence is in their favor, which distracts you from the over-all fact that your cereal tastes worse and is less healthy."

After Haidt's piece was published, the Google Doc—"Social Media and Political Dysfunction: A Collaborative Review"—was made available to the public. Comments piled up, and a new section was added, at the end, to include a miscellany of Twitter threads and Substack essays that appeared in response to Haidt's interpretation of the evidence. Some colleagues and kibbitzers agreed with Haidt. But others, though they might have shared his basic intuition that *something* in our experience of social media was amiss, drew upon the same data set to reach less definitive conclusions, or even mildly contradictory ones. Even

after the initial flurry of responses to Haidt's article disappeared into social-media memory, the document, insofar as it captured the state of the social-media debate, remained a lively artifact.

Near the end of the collaborative project's introduction, the authors warn, "We caution readers not to simply add up the number of studies on each side and declare one side the winner." The document runs to more than a hundred and fifty pages, and for each question there are affirmative and dissenting studies, as well as some that indicate mixed results. According to one paper, "Political expressions on social media and the online forum were found to (a) reinforce the expressers' partisan thought process and (b) harden their pre-existing political preferences," but, according to another, which used data collected during the 2016 election, "Over the course of the campaign, we found media use and attitudes remained relatively stable. Our results also showed that Facebook news use was related to modest over-time spiral of depolarization. Furthermore, we found that people who use Facebook for news were more likely to view both pro- and counter-attitudinal news in each wave. Our results indicated that counter-attitudinal exposure increased over time, which resulted in depolarization." If results like these seem incompatible, a perplexed reader is given recourse to a study that says, "Our findings indicate that political polarization on social media cannot be conceptualized as a unified phenomenon, as there are significant cross-platform differences."

Interested in echo chambers? "Our results show that the aggregation of users in homophilic clusters dominate online interactions on Facebook and Twitter," which seems convincing—except that, as another team has it, "We do not find evidence supporting a strong characterization of 'echo chambers' in which the majority of people's sources of news are mutually exclusive and from opposite poles." By the end of the file, the vaguely patronizing top-line recommendation against simple summation begins to make more sense. A document that originated as a bulwark

against confirmation bias could, as it turned out, just as easily function as a kind of generative device to support anybody's pet conviction. The only sane response, it seemed, was simply to throw one's hands in the air.

When I spoke to some of the researchers whose work had been included, I found a combination of broad, visceral unease with the current situation—with the banefulness of harassment and trolling; with the opacity of the platforms; with, well, the widespread presentiment that of course social media is in many ways bad—and a contrastive sense that it might not be *catastrophically* bad in some of the specific ways that many of us have come to take for granted as true. This was not mere contrarianism, and there was no trace of gleeful mythbusting; the issue was important enough to get right. When I told Bail that the upshot seemed to me to be that exactly nothing was unambiguously clear, he suggested that there was at least some firm ground. He sounded a bit less apocalyptic than Haidt.

"A lot of the stories out there are just wrong," he told me. "The political echo chamber has been massively overstated. Maybe it's three to five per cent of people who are properly in an echo chamber." Echo chambers, as hotboxes of confirmation bias, are counterproductive for democracy. But research indicates that most of us are actually exposed to a wider range of views on social media than we are in real life, where our social networks—in the original use of the term—are rarely heterogeneous. (Haidt told me that this was an issue on which the Google Doc changed his mind; he became convinced that echo chambers probably aren't as widespread a problem as he'd once imagined.) And too much of a focus on our intuitions about social media's echo-chamber effect could obscure the relevant counterfactual: a conservative might abandon Twitter only to watch more Fox News. "Stepping outside your echo chamber is supposed to make you moderate, but maybe it makes you more extreme," Bail said. The research is inchoate and ongoing, and it's difficult to say anything on the topic with absolute certainty. But this was, in part, Bail's point: we ought to be less sure about the particular impacts of social media.

Bail went on, "The second story is foreign misinformation." It's not that misinformation doesn't exist, or that it hasn't had indirect effects, especially when it creates perverse incentives for the mainstream media to cover stories circulating online. Haidt also draws convincingly upon the work of Renée DiResta, the research manager at the Stanford Internet Observatory, to sketch out a potential future in which the work of shitposting has been outsourced to artificial intelligence, further polluting the informational environment. But, at least so far, very few Americans seem to suffer from consistent exposure to fake news —"probably less than two per cent of Twitter users, maybe fewer now, and for those who were it didn't change their opinions," Bail said. This was probably because the people likeliest to consume such spectacles were the sort of people primed to believe them in the first place. "In fact," he said, "echo chambers might have done something to quarantine that misinformation."

The final story that Bail wanted to discuss was the "proverbial rabbit hole, the path to algorithmic radicalization," by which YouTube might serve a viewer increasingly extreme videos. There is some anecdotal evidence to suggest that this does happen, at least on occasion, and such anecdotes are alarming to hear. But a new working paper led by Brendan Nyhan, a political scientist at Dartmouth, found that almost all extremist content is either consumed by subscribers to the relevant channels—a sign of actual demand rather than manipulation or preference falsification—or encountered via links from external sites. It's easy to see why we might prefer if this were not the case: algorithmic radicalization is presumably a simpler problem to solve than the fact that there are people who deliberately seek out vile content. "These are the three stories—echo chambers, foreign influence campaigns, and radicalizing recommendation algorithms—but, when you look at the literature, they've all been overstated." He thought that these findings were crucial for us to assimilate, if only to help us understand that our problems may lie beyond technocratic tinkering. He explained, "Part of my interest in getting this research out there is to demonstrate that everybody is

waiting for an Elon Musk to ride in and save us with an algorithm"—or, presumably, the reverse—"and it's just not going to happen."

When I spoke with Nyhan, he told me much the same thing: "The most credible research is way out of line with the takes." He noted, of extremist content and misinformation, that reliable research that "measures exposure to these things finds that the people consuming this content are small minorities who have extreme views already." The problem with the bulk of the earlier research, Nyhan told me, is that it's almost all correlational. "Many of these studies will find polarization on social media," he said. "But that might just be the society we live in reflected on social media!" He hastened to add, "Not that this is untroubling, and none of this is to let these companies, which are exercising a lot of power with very little scrutiny, off the hook. But a lot of the criticisms of them are very poorly founded. . . . The expansion of Internet access coincides with fifteen other trends over time, and separating them is very difficult. The lack of good data is a huge problem insofar as it lets people project their own fears into this area." He told me, "It's hard to weigh in on the side of 'We don't know, the evidence is weak,' because those points are always going to be drowned out in our discourse. But these arguments are systematically underprovided in the public domain."

In his *Atlantic* article, Haidt leans on a working paper by two social scientists, Philipp Lorenz-Spreen and Lisa Oswald, who took on a comprehensive meta-analysis of about five hundred papers and concluded that "the large majority of reported associations between digital media use and trust appear to be detrimental for democracy." Haidt writes, "The literature is complex—some studies show benefits, particularly in less developed democracies—but the review found that, on balance, social media amplifies political polarization; foments populism, especially right-wing populism; and is associated with the spread of misinformation." Nyhan was less convinced that the meta-analysis supported such categorical verdicts, especially once you bracketed the kinds of correlational findings that might simply

mirror social and political dynamics. He told me, "If you look at their summary of studies that allow for causal inferences—it's very mixed."

As for the studies Nyhan considered most methodologically sound, he pointed to a 2020 article called "The Welfare Effects of Social Media," by Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. For four weeks prior to the 2018 midterm elections, the authors randomly divided a group of volunteers into two cohorts—one that continued to use Facebook as usual, and another that was paid to deactivate their accounts for that period. They found that deactivation "(i) reduced online activity, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in post-experiment Facebook use." But Gentzkow reminded me that his conclusions, including that Facebook may slightly increase polarization, had to be heavily qualified: "From other kinds of evidence, I think there's reason to think social media is not the main driver of increasing polarization over the long haul in the United States."

In the book "Why We're Polarized," for example, Ezra Klein invokes the work of such scholars as Lilliana Mason to argue that the roots of polarization might be found in, among other factors, the political realignment and nationalization that began in the sixties, and were then sacralized, on the right, by the rise of talk radio and cable news. These dynamics have served to flatten our political identities, weakening our ability or inclination to find compromise. Insofar as some forms of social media encourage the hardening of connections between our identities and a narrow set of opinions, we might increasingly self-select into mutually incomprehensible and hostile groups; Haidt plausibly suggests that these processes are accelerated by the coalescence of social-media tribes around figures of fearful online charisma. "Social media might be more of an amplifier of other things going on rather than a major driver independently," Gentzkow argued. "I think it takes some gymnastics to tell a story where it's all primarily driven by social media, especially when you're looking at different countries, and across different groups."

Another study, led by Nejla Asimovic and Joshua Tucker, replicated Gentzkow's approach in Bosnia and Herzegovina, and they found almost precisely the opposite results: the people who stayed on Facebook were, by the end of the study, *more* positively disposed to their historic out-groups. The authors' interpretation was that ethnic groups have so little contact in Bosnia that, for some people, social media is essentially the only place where they can form positive images of one another. "To have a replication and have the signs flip like that, it's pretty stunning," Bail told me. "It's a different conversation in every part of the world."

Nyhan argued that, at least in wealthy Western countries, we might be too heavily discounting the degree to which platforms have responded to criticism: "Everyone is still operating under the view that algorithms simply maximize engagement in a short-term way" with minimal attention to potential externalities. "That might've been true when Zuckerberg had seven people working for him, but there are a lot of considerations that go into these rankings now." He added, "There's some evidence that, with reverse-chronological feeds"—streams of unwashed content, which some critics argue are less manipulative than algorithmic curation—"people get exposed to *more* low-quality content, so it's another case where a very simple notion of 'algorithms are bad' doesn't stand up to scrutiny. It doesn't mean they're good, it's just that we don't know."

Bail told me that, over all, he was less confident than Haidt that the available evidence lines up clearly against the platforms. "Maybe there's a slight majority of studies that say that social media is a net negative, at least in the West, and maybe it's doing some good in the rest of the world." But, he noted, "Jon will say that science has this expectation of rigor that can't keep up with the need in the real world—that even if we don't have the definitive study that creates the historical counterfactual that Facebook is largely responsible for polarization in the U.S., there's still a lot pointing in that direction, and I think that's a fair point." He paused. "It can't all be randomized control trials."

Haidt comes across in conversation as searching and sincere, and, during our exchange, he paused several times to suggest that I include a quote from John Stuart Mill on the importance of good-faith debate to moral progress. In that spirit, I asked him what he thought of the argument, elaborated by some of Haidt's critics, that the problems he described are fundamentally political, social, and economic, and that to blame social media is to search for lost keys under the streetlamp, where the light is better. He agreed that this was the steelman opponent: there were predecessors for cancel culture in de Tocqueville, and anxiety about new media that went back to the time of the printing press. "This is a perfectly reasonable hypothesis, and it's absolutely up to the prosecution—people like me—to argue that, no, this time it's different. But it's a civil case! The evidential standard is not 'beyond a reasonable doubt,' as in a criminal case. It's just a preponderance of the evidence."

The way scholars weigh the testimony is subject to their disciplinary orientations. Economists and political scientists tend to believe that you can't even begin to talk about causal dynamics without a randomized controlled trial, whereas sociologists and psychologists are more comfortable drawing inferences on a correlational basis. Haidt believes that conditions are too dire to take the hardheaded, no-reasonable-doubt view. "The preponderance of the evidence is what we use in public health. If there's an epidemic—when COVID started, suppose all the scientists had said, 'No, we gotta be so certain before you do anything'? We have to think about what's actually happening, what's likeliest to pay off." He continued, "We have the largest epidemic ever of teen mental health, and there is no other explanation," he said. "It is a raging public-health epidemic, and the kids themselves say Instagram did it, and we have some evidence, so is it appropriate to say, 'Nah, you haven't proven it'?"

This was his attitude across the board. He argued that social media seemed to aggrandize inflammatory posts and to be correlated with a rise in violence; even if

only small groups were exposed to fake news, such beliefs might still proliferate in ways that were hard to measure. "In the post-Babel era, what matters is not the average but the dynamics, the contagion, the exponential amplification," he said. "Small things can grow very quickly, so arguments that Russian disinformation didn't matter are like COVID arguments that people coming in from China didn't have contact with a lot of people." Given the transformative effects of social media, Haidt insisted, it was important to act now, even in the absence of dispositive evidence. "Academic debates play out over decades and are often never resolved, whereas the social-media environment changes year by year," he said. "We don't have the luxury of waiting around five or ten years for literature reviews."

Haidt could be accused of question-begging—of assuming the existence of a crisis that the research might or might not ultimately underwrite. Still, the gap between the two sides in this case might not be quite as wide as Haidt thinks. Skeptics of his strongest claims are not saying that there's no *there* there. Just because the average YouTube user is unlikely to be led to Stormfront videos, Nyhan told me, doesn't mean we shouldn't worry that some people are watching Stormfront videos; just because echo chambers and foreign misinformation seem to have had effects only at the margins, Gentzkow said, doesn't mean they're entirely irrelevant. "There are many questions here where the thing we as researchers are interested in is how social media affects the *average* person," Gentzkow told me. "There's a different set of questions where all you need is a small number of people to change —questions about ethnic violence in Bangladesh or Sri Lanka, people on YouTube mobilized to do mass shootings. Much of the evidence broadly makes me skeptical that the average effects are as big as the public discussion thinks they are, but I also think there are cases where a small number of people with very extreme views are able to find each other and connect and act." He added, "That's where many of the things I'd be most concerned about lie."

The same might be said about any phenomenon where the base rate is very low but the stakes are very high, such as teen suicide. "It's another case where those rare edge cases in terms of total social harm may be enormous. You don't need

many teen-age kids to decide to kill themselves or have serious mental-health outcomes in order for the social harm to be really big." He added, "Almost none of this work is able to get at those edge-case effects, and we have to be careful that if we do establish that the average effect of something is zero, or small, that it doesn't mean we shouldn't be worried about it—because we might be missing those extremes." Jaime Settle, a scholar of political behavior at the College of William & Mary and the author of the book "Frenemies: How Social Media Polarizes America," noted that Haidt is "farther along the spectrum of what most academics who study this stuff are going to say we have strong evidence for." But she understood his impulse: "We do have serious problems, and I'm glad Jon wrote the piece, and down the road I wouldn't be surprised if we got a fuller handle on the role of social media in all of this—there are definitely ways in which social media has changed our politics for the worse."

It's tempting to sidestep the question of diagnosis entirely, and to evaluate Haidt's essay not on the basis of predictive accuracy—whether social media will lead to the destruction of American democracy—but as a set of proposals for what we might do better. If he is wrong, how much damage are his prescriptions likely to do? Haidt, to his great credit, does not indulge in any wishful thinking, and if his diagnosis is largely technological his prescriptions are sociopolitical. Two of his three major suggestions seem useful and have nothing to do with social media: he thinks that we should end closed primaries and that children should be given wide latitude for unsupervised play. His recommendations for social-media reform are, for the most part, uncontroversial: he believes that preteens shouldn't be on Instagram and that platforms should share their data with outside researchers—proposals that are both likely to be beneficial and not very costly.

It remains possible, however, that the true costs of social-media anxieties are harder to tabulate. Gentzkow told me that, for the period between 2016 and 2020, the direct effects of misinformation were difficult to discern. "But it might have

had a much larger effect because we got so worried about it—a broader impact on trust," he said. "Even if not that many people were exposed, the narrative that the world is full of fake news, and you can't trust anything, and other people are being misled about it—well, that might have had a bigger impact than the content itself." Nyhan had a similar reaction. "There are genuine questions that are really important, but there's a kind of opportunity cost that is missed here. There's so much focus on sweeping claims that aren't actionable, or unfounded claims we can contradict with data, that are crowding out the harms we *can* demonstrate, and the things we *can* test, that could make social media better." He added, "We're years into this, and we're still having an uninformed conversation about social media. It's totally wild."

New Yorker Favorites

- What the midterms mean for Donald Trump.
- The doxing of Jane Doe and the fallout from the "Cornell Seven" rape case.
- "She Made Me Do It": a famed equestrian instructor shot his most annoying student.
- Parents disagreed on whether to vaccinate their children. Then the kids got sick.
- Play a new game of Laugh Lines every day of the week!

Want better search results? Make *The New Yorker* a preferred source in Google.

Sign up for our daily newsletter to receive the best stories from *The New Yorker*.



Gideon Lewis-Kraus is a staff writer at The New Yorker covering technology, academia, and books, among other topics. He is the author of the memoir "A Sense of Direction."

More: Social Media Politics Psychology Polarization Political Science Misinformation

The Goings On Newsletter

What we're watching, listening to, and doing this week in N.Y.C, and beyond

SIGN UP

By signing up, you agree to our user agreement (including class action waiver and arbitration provisions), and acknowledge our privacy policy.

READ MORE

THE LEDE

Is Trump's Latest Press Attack Just Another Distraction?

After CNN, MS NOW, and *Politico* were barred from the White House, the entire TV press pool announced a boycott of the President.

By Jon Allsop

THE LEDE

The Humanism Driving the Anti-A.I. Movement

At the Pro-Human Assembly in Washington, D.C., attendees talked less about jobs and policy, and more about their philosophy of life.

By Becca Rothfeld

ON TELEVISION

The Bari Weiss Effect on "60 Minutes"

CBS News' editor-in-chief has pledged to "speak for the seventy-five per cent" of people on the center right and the center left; the reality is different.

By Vinson Cunningham

THE LEDE

There Is No A.I. "Race"

We're constantly told that Chinese advances would have catastrophic consequences. But the first to innovate is rarely the one who wins.

By Chang Che

INFINITE SCROLL

Caleb Hammer's Humiliation Economy

The host of "Financial Audit," a popular YouTube series, claims that his goal is to make his viewers more financially literate, with his down-and-out interviewees serving as extreme test cases for what not to do.

By Brady Brickner-Wood

ANNALS OF MONEY

The Battle to Pass a Billionaire Tax

Most Democratic voters say that they favor a wealth tax, but a California ballot initiative has plenty of liberal opponents.

By Nicholas Lemann

THE LEDE

The Toxic Logic Behind the Shaming of Cornell's Jane Doe

One of the biggest failures of the post-#MeToo moment is our collective inability to provide a confident, humane account of sex as a basic good.

By Anahid Nersessian

INFINITE SCROLL

How Lists Became Our Greatest Defense Against Endless Slop

A form that was once emblematic of the internet's frivolity has become our preferred way of expressing taste, with the New York *Times* and Letterboxd users alike clamoring to define the canon.

By Brady Brickner-Wood

THE FRONT ROW

"The Social Reckoning" Leaves Out the Human Element

In Aaron Sorkin's overwritten sequel to "The Social Network," inspired performances by Mikey Madison and Jeremy Allen White can't bring the main characters to life.

By Richard Brody

INFINITE SCROLL

Hollywood's Tech-Founder Fixation Takes a Dark Turn

Tech-entrepreneur bio-pics used to give us flawed men who built great products. But the latest wave of movies, from "The Social Reckoning" to "Artificial," recast Silicon Valley titans as unambiguous villains, and suggest that the products themselves are the threat.

By Brady Brickner-Wood

THE POLITICAL SCENE

The Battle for the Post-Trump Right

For decades, the Heritage Foundation was a conservative powerhouse. Can it stay relevant in the influencer age?

By Emma Green

FAULT LINES

Individual Choices Won't Fix America's Public Schools

Nikole Hannah-Jones sent her daughter to a struggling local school and has come to regret it. But personalizing the problem is the wrong approach.

By Jay Caspian Kang

Privacy Information