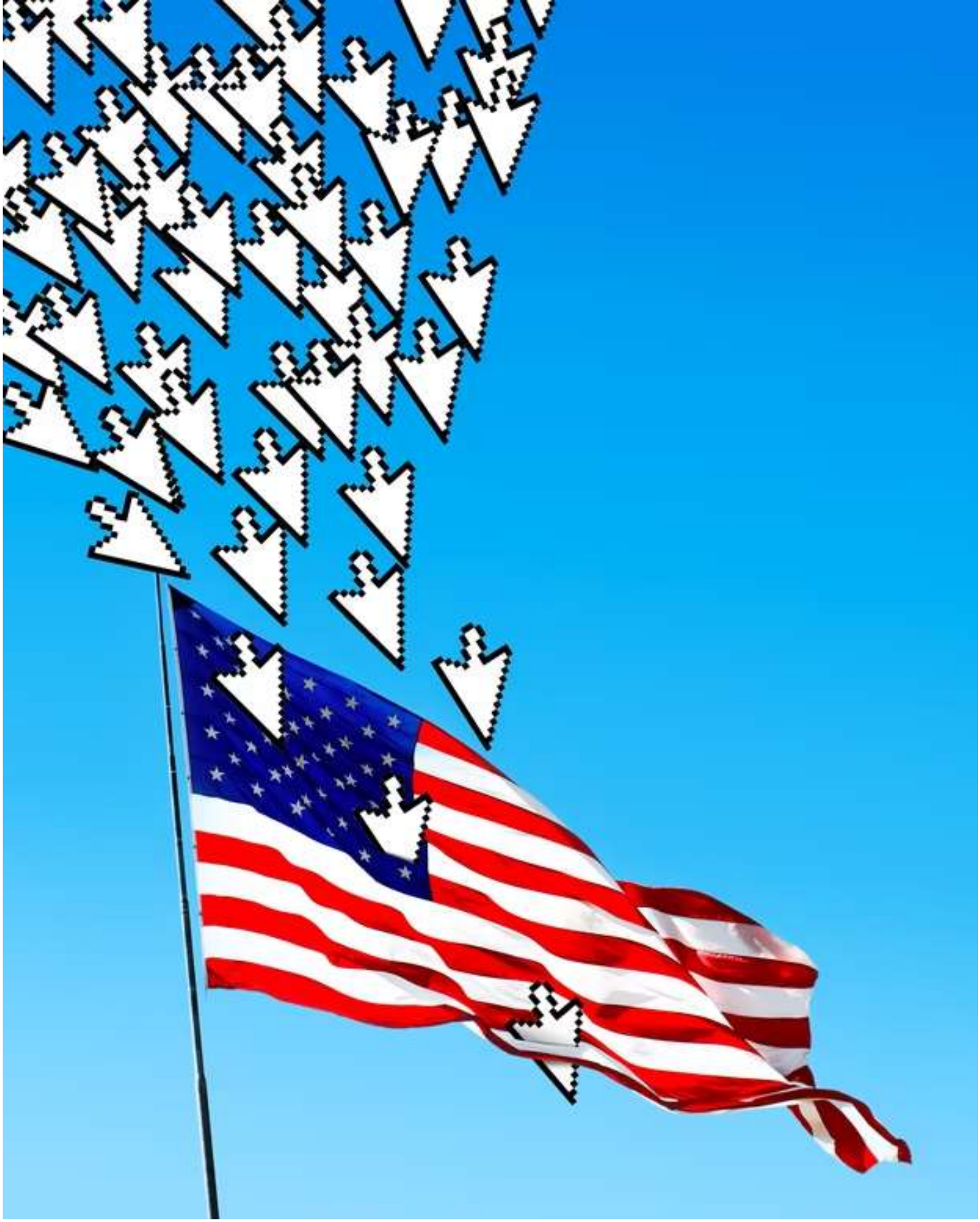
IDEAS

YES, SOCIAL MEDIA REALLY IS UNDERMINING DEMOCRACY

Despite what Meta has to say.

By Jonathan Haidt



Katie Martin / The Atlantic; Getty

JULY 28, 2022, 6 AM ET

SHARE

About the author: *Jonathan Haidt is a social psychologist at the New York University Stern School of Business. He is the co-author of* The Coddling of the American Mind, *which originated as a September 2015* Atlantic *story, and he is currently writing* Life After Babel: Adapting to a World We Can No Longer Share.

WITHIN THE PAST 15 years, social media has insinuated itself into American life more deeply than food-delivery apps into our diets and microplastics into our bloodstreams. Look at stories about conflict, and it's often lurking in the background. Recent articles on the rising dysfunction within progressive organizations point to the role of Twitter, Slack, and other platforms in prompting "endless and sprawling internal microbattles," as *The Intercept*'s Ryan Grim put it, referring to the ACLU. At a far higher level of conflict, the congressional hearings about the January 6 insurrection show us how Donald Trump's tweets summoned the mob to Washington and aimed it at the vice president. Far-right groups then used a variety of platforms to coordinate and carry out the attack.

Social media has changed life in America in a thousand ways, and nearly two out of three Americans now believe that these changes are for the worse. But academic researchers have not yet reached a consensus that social media is harmful. That's been a boon to social-media companies such as Meta, which argues, as did tobacco companies, that the science is not "settled."

The lack of consensus leaves open the possibility that social media may not be very harmful. Perhaps we've fallen prey to yet another moral panic about a new technology and, as with television, we'll worry about it less after a few decades of conflicting studies. A different possibility is that social media is quite harmful but is changing too quickly for social scientists to capture its effects. The research community is built on a quasi-moral norm of skepticism: We begin by assuming the null hypothesis (in this case, that social media is *not* harmful), and we require researchers to show strong, statistically significant evidence in order to publish their findings. This takes time—a couple of years, typically, to conduct and publish a study; five or more years before review papers and meta-analyses come out; sometimes decades before scholars reach

agreement. Social-media platforms, meanwhile, can change dramatically in just a few years.

So even if social media really did begin to undermine democracy (and institutional trust and teen mental health) in the early 2010s, we should not expect social science to "settle" the matter until the 2030s. By then, the effects of social media will be radically different, and the harms done in earlier decades may be irreversible.

Let me back up. This spring, *The Atlantic* published my essay "Why the Past 10 Years of American Life Have Been Uniquely Stupid," in which I argued that the best way to understand the chaos and fragmentation of American society is to see ourselves as citizens of Babel in the days after God rendered them unable to understand one another.

I showed how a few small changes to the architecture of social-media platforms, implemented from 2009 to 2012, increased the virality of posts on those platforms, which then changed the nature of social relationships. People could spread rumors and half-truths more quickly, and they could more readily sort themselves into homogenous tribes. Even more important, in my view, was that social-media platforms such as Twitter and Facebook could now be used more easily by anyone to attack anyone. It was as if the platforms had passed out a billion little dart guns, and although most users didn't want to shoot anyone, three kinds of people began darting others with abandon: the far right, the far left, and trolls.

Jonathan Haidt and Tobias Rose-Stockwell: The dark psychology of social networks

All of these groups were suddenly given the power to dominate conversations and intimidate dissenters into silence. A fourth group—Russian agents—also got a boost, though they didn't need to attack people directly. Their long-running project, which ramped up online in 2013, was to fabricate, exaggerate, or

RECOMMENDED READING

Your Pillows Might Be Killing Your Neck

OLGA KHAZAN

How to Find the Secret to Meaningful Work

REBECCA RASHID AND ARTHUR C. BROOKS

simply promote stories that would increase Americans' hatred of one another and distrust of their institutions.

The Tyranny of Workplace Food-Shamers

AMANDA MULL

The essay proved to be surprisingly uncontroversial—or, at least, hardly anyone attacked me on social media. But a few responses were published, including one from Meta (formerly Facebook), which pointed to studies it said contradicted my argument. There was also an essay in *The New Yorker* by Gideon Lewis-Kraus, who interviewed me and other scholars who study politics and social media. He argued that social media might well be harmful to democracies, but the research literature is too muddy and contradictory to support firm conclusions.

So was my diagnosis correct, or are concerns about social media overblown? It's a crucial question for the future of our society. As I argued in my essay, critics make us smarter. I'm grateful, therefore, to Meta and the researchers interviewed by Lewis-Kraus for helping me sharpen and extend my argument in three ways.

Are Democracies Becoming More Polarized and Less Healthy?

My essay laid out a wide array of harms that social media has inflicted on society. Political polarization is just one of them, but it is central to the story of rising democratic dysfunction.

Meta questioned whether social media should be blamed for increased polarization. In response to my essay, Meta's head of research, Pratiti Raychoudhury, pointed to a study by Levi Boxell, Matthew Gentzkow, and Jesse Shapiro that looked at trends in 12 countries and found, she said, "that in some countries polarization was on the rise before Facebook even existed, and in others it has been decreasing while internet and Facebook use increased." In a recent interview with the podcaster Lex Fridman, Mark Zuckerberg cited this same study in support of a more audacious claim: "Most of the academic studies that I've seen actually show that social-media use is correlated with lower polarization."

Does that study really let social media off the hook? It plotted political polarization based on survey responses in 12 countries, most with data stretching back to the 1970s, and then drew straight lines that best fit the data points over several decades.

It's true that, while some lines sloped upward (meaning that polarization increased across the period as a whole), others sloped downward. But my argument wasn't about the past 50 years. It was about a phase change that happened *in the early 2010s*, after Facebook and Twitter changed their architecture to enable hyper-virality.

I emailed Gentzkow to ask whether he could put a "hinge" in the graphs in the early 2010s, to see if the trends in polarization changed direction or accelerated in the past decade. He replied that there was not enough data after 2010 to make such an analysis reliable. He also noted that Meta's response essay had failed to cite a 2020 article in which he and three colleagues found that randomly assigning participants to deactivate Facebook for the four weeks before the 2018 U.S. midterm elections reduced polarization.

Adrienne LaFrance: 'History will not judge us kindly'

Meta's response motivated me to look for additional publications to evaluate what had happened to democracies in the 2010s. I discovered four. One of them found no overall trend in polarization, but like the study by Boxell, Gentzkow, and Shapiro, it had few data points after 2015. The other three had data through 2020, and *all three* reported substantial increases in polarization and/or declines in the number or quality of democracies around the world.

One of them, a 2022 report from the Varieties of Democracy (V-Dem) Institute, found that "liberal democracies peaked in 2012 with 42 countries and are now down to the lowest levels in over 25 years." It summarized the transformations of global democracy over the past 10 years in stark terms:

> Just ten years ago the world looked very different from today. In 2011, there were more countries improving than declining on every aspect of democracy. By 2021 the world has been turned on its head: there are more countries declining than advancing on nearly all democratic aspects captured by V-Dem measures.

The report also notes that "toxic polarization"—signaled by declining "respect for counter-arguments and associated aspects of the deliberative component of democracy"—grew more severe in at least 32 countries.

A paper published one week after my *Atlantic* essay, by Yunus E. Orhan, found a global spike in democratic "backsliding" since 2008, and linked it to affective polarization, or animosity toward the other side. When affective polarization is high, partisans tolerate antidemocratic behavior by politicians on their own side—such as the January 6 attack on the U.S. Capitol.

And finally, the Economist Intelligence Unit reported a global decline in various democratic measures starting after 2015, according to its Democracy Index.

These three studies cannot prove that social media *caused* the global decline, but—contra Meta and Zuckerberg—they show a global trend toward polarization in the previous decade, the one in which the world embraced social media.

Has Social Media Created Harmful Echo Chambers?

So why did democracies weaken in the 2010s? How might social media have made them more fragmented and less stable? One popular argument contends that social media sorts users into echo chambers—closed communities of like-minded people. Lack of contact with people who hold different viewpoints allows a sort of tribal groupthink to take hold, reducing the quality of everyone's thinking and the prospects for compromise that are essential in a democratic system.

According to Meta, however, "More and more research discredits the idea that social media algorithms create an echo chamber." It points to two sources to back up that claim, but many studies show evidence that social media *does* in fact create echo chambers. Because conflicting studies are common in social-science research, I created a "collaborative review" document last year with Chris Bail, a sociologist at Duke University who studies social media. It's a public Google doc in which we organize the abstracts of all the studies we can find about social media's impact on democracy, and then we invite other experts to add studies, comments, and criticisms. We cover research on seven different questions, including whether social media promotes echo chambers. After spending time in the document, Lewis-Kraus wrote in *The New*

Yorker: "The upshot seemed to me to be that exactly nothing was unambiguously clear."

He is certainly right that nothing is unambiguous. But as I have learned from curating three such documents, researchers often reach opposing conclusions because they have "operationalized" the question differently. That is, they have chosen different ways to turn an abstract question (about the prevalence of echo chambers, say) into something concrete and measurable. For example, researchers who choose to measure echo chambers by looking at the diversity of people's news consumption typically find little evidence that they exist at all. Even partisans end up being exposed to news stories and videos from the other side. Both of the sources that Raychoudhury cited in her defense of Meta mention this idea.

Derek Thompson: Social media is attention alcohol

But researchers who measure echo chambers by looking at social relationships and networks usually find evidence of "homophily"—that is, people tend to engage with others who are similar to themselves. One study of politically engaged Twitter users, for example, found that they "are disproportionately exposed to like-minded information and that information reaches like-minded users more quickly."

So should we throw up our hands and say that the findings are irreconcilable? No, we should integrate them, as the sociologist Zeynep Tufekci did in a 2018 essay. Coming across contrary viewpoints on social media, she wrote, is "not like reading them in a newspaper while sitting alone." Rather, she said, "it's like hearing them from the opposing team while sitting with our fellow fans in a football stadium … We bond with our team by yelling at the fans of the other one." Mere exposure to different sources of news doesn't automatically break open echo chambers; in fact, it can reinforce them.

These closely bonded groupings can have profound political ramifications, as a couple of my critics in the *New Yorker* article acknowledged. A major feature of the post-Babel world is that the extremes are now far louder and more influential than before. They may also become more violent. Recent research by Morteza Dehghani and his colleagues at the University of Southern California shows that people are more willing

to commit violence when they are immersed in a community they perceive to be morally homogeneous.

This finding seems to be borne out by a statement from the 18-year-old man who recently killed 10 Black Americans at a supermarket in Buffalo. In the Q&A portion of the manifesto attributed to him, he wrote:

> **Where did you get your current beliefs?**
>
> Mostly from the internet. There was little to no influence on my personal beliefs by people I met in person.

The killer goes on to claim that he had read information "from all ideologies," but I find it unlikely that he consumed a balanced informational diet, or, more important, that he hung out online with ideologically diverse users. The fact that he livestreamed his shooting tells us he assumed that his community shared his warped worldview. He could not have found such an extreme yet homogeneous group in his small town 200 miles from Buffalo. But thanks to social media, he found an international fellowship of extreme racists who jointly worshipped past mass murderers and from whom he copied sections of his manifesto.

Is Social Media the *Primary* Villain in This Story?

In her response to my essay, Raychoudhury did not deny that Meta bore any blame. Rather, her defense was two-pronged, arguing that the research is not yet definitive, and that, in any case, we should be focusing on mainstream media as the primary cause of harm.

Raychoudhury pointed to a study on the role of cable TV and mainstream media as major drivers of partisanship. She is correct to do so: The American culture war has roots going back to the turmoil of the 1960s, which activated evangelicals and other conservatives in the '70s. Social media (which arrived around 2004 and became truly pernicious, I argue, only after 2009) is indeed a more recent player in this phenomenon.

In my essay, I included a paragraph on this backstory, noting the role of Fox News and the radicalizing Republican Party of the '90s, but I should have said more. The story of polarization is complex, and political scientists cite a variety of contributing factors, including the growing politicization of the urban-rural divide; rising immigration; the increasing power of big and very partisan donors; the loss of a common enemy when the Soviet Union collapsed; and the loss of the "Greatest Generation," which had an ethos of service forged in the crisis of the Second World War. And although polarization rose rapidly in the 2010s, the rise began in the '90s, so I cannot pin the *majority* of the rise on social media.

But my essay wasn't primarily about ordinary polarization. I was trying to explain a new dynamic that emerged in the 2010s: *the fear of one another*, even—and perhaps especially—within groups that share political or cultural affinities. This fear has created a whole new set of social and political problems.

The loss of a common enemy and those other trends with roots in the 20th century can help explain America's ever nastier cross-party relationships, but they can't explain why so many college students and professors suddenly began to express more fear, and engage in more self-censorship, around 2015. These mostly left-leaning people weren't worried about the "other side"; they were afraid of a small number of students who were *further* to the left, and who enthusiastically hunted for verbal transgressions and used social media to publicly shame offenders.

A few years later, that same fearful dynamic spread to newsrooms, companies, nonprofit organizations, and many other parts of society. The culture war had been running for two or three decades by then, but it changed in the mid-2010s when ordinary people with little to no public profile suddenly became the targets of social-media mobs. Consider the famous 2013 case of Justine Sacco, who tweeted an insensitive joke about her trip to South Africa just before boarding her flight in London and became an international villain by the time she landed in Cape Town. She was fired the next day. Or consider the the far right's penchant for using social media to publicize the names and photographs of largely unknown local election officials, health officials, and school-board members who refuse to bow to political pressure, and who are then subjected to waves of vitriol, including threats of violence to themselves and their children, simply for doing their jobs. These phenomena, now

common to the culture, could not have happened before the advent of hyper-viral social media in 2009.

Matthew Hindman, Nathaniel Lubin, and Trevor Davis: Facebook has a superuser-supremacy problem

This fear of getting shamed, reported, doxxed, fired, or physically attacked is responsible for the self-censorship and silencing of dissent that were the main focus of my essay. When dissent within any group or institution is stifled, the group will become less perceptive, nimble, and effective over time.

Social media may not be the *primary* cause of polarization, but it is an important cause, and one we can do something about. I believe it is also the primary cause of the epidemic of structural stupidity, as I called it, that has recently afflicted many of America's key institutions.

What Can We Do to Make Things Better?

My essay presented a series of structural solutions that would allow us to repair some of the damage that social media has caused to our key democratic and epistemic institutions. I proposed three imperatives: (1) harden democratic institutions so that they can withstand chronic anger and mistrust, (2) reform social media so that it becomes less socially corrosive, and (3) better prepare the next generation for democratic citizenship in this new age.

I believe that we should begin implementing these reforms now, even if the science is not yet "settled." *Beyond a reasonable doubt* is the appropriate standard of evidence for reviewers guarding admission to a scientific journal, or for jurors establishing guilt in a criminal trial. It is too high a bar for questions about public health or threats to the body politic. A more appropriate standard is the one used in civil trials: the *preponderance* of evidence. Is social media *probably* damaging American democracy via at least one of the seven pathways analyzed in our collaborative-review document, or *probably not*? I urge readers to examine the document themselves. I also urge the social-science community to find quicker ways to study potential threats such as social media, where platforms and their effects change rapidly. Our motto should be "Move

fast and test things." Collaborative-review documents are one way to speed up the process by which scholars find and respond to one another's work.

Beyond these structural solutions, I considered adding a short section to the article on what each of us can do as individuals, but it sounded a bit too preachy, so I cut it. I now regret that decision. I should have noted that all of us, as individuals, can be part of the solution by choosing to act with courage, moderation, and compassion. It takes a great deal of resolve to speak publicly or stand your ground when a barrage of snide, disparaging, and otherwise hostile comments is coming at you and nobody rises to your defense (out of fear of getting attacked themselves).

Read: How to fix Twitter—and all of social media

Fortunately, social media does not usually reflect real life, something that more people are beginning to understand. A few years ago, I heard an insight from an older business executive. He noted that before social media, if he received a dozen angry letters or emails from customers, they spurred him to action because he assumed that there must be a thousand other disgruntled customers who didn't bother to write. But now, if a thousand people like an angry tweet or Facebook post about his company, he assumes that there must be a dozen people who are really upset.

Seeing that social-media outrage is transient and performative should make it easier to withstand, whether you are the president of a university or a parent speaking at a school-board meeting. We can all do more to offer honest dissent and support the dissenters within institutions that have become structurally stupid. We can all get better at listening with an open mind and speaking in order to engage another human being rather than impress an audience. Teaching these skills to our children and our students is crucial, because they are the generation who will have to reinvent deliberative democracy and Tocqueville's "art of association" for the digital age.

We must act with compassion too. The fear and cruelty of the post-Babel era are a result of its tendency to reward public displays of aggression. Social media has put us all in the middle of a Roman coliseum, and many in the audience want to see conflict and blood. But once we realize that we are the gladiators—tricked into combat so that we might generate "content," "engagement," and revenue—we can refuse to fight. We can be more understanding toward our fellow citizens, seeing that we are all being

driven mad by companies that use largely the same set of psychological tricks. We can forswear public conflict and use social media to serve our own purposes, which for most people will mean more private communication and fewer public performances.

The post-Babel world will not be rebuilt by today's technology companies. That work will be left to citizens who understand the forces that brought us to the verge of self-destruction, and who develop the new habits, virtues, technologies, and shared narratives that will allow us to reap the benefits of living and working together in peace.