

JULY 23, 2019 | NUMBER 876

## Challenging the Social Media Moral Panic

### Preserving Free Expression under Hypertransparency

BY MILTON MUELLER

#### EXECUTIVE SUMMARY

**S**ocial media are now widely criticized after enjoying a long period of public approbation. The kinds of human activities that are coordinated through social media, good as well as bad, have always existed. However, these activities were not visible or accessible to the whole of society. As conversation, socialization, and commerce are aggregated into large-scale, public commercial platforms, they become highly visible to the public and generate storable, searchable records. Social media make human interactions hypertransparent and displace the responsibility for societal acts from the perpetrators to the platform that makes them visible.

This hypertransparency is fostering a moral panic around social media. Internet platforms, like earlier new media technologies such as TV and radio, now stand accused of a stunning array of evils: addiction, fostering terrorism and extremism, facilitating ethnic cleansing, and even the destruction of democracy. The social-psychological dynamics of hypertransparency lend themselves to the conclusion that social media cause the problems they reveal and that society would be

improved by regulating the intermediaries that facilitate unwanted activities.

This moral panic should give way to calmer reflection. There needs to be a clear articulation of the tremendous value of social media platforms based on their ability to match seekers and providers of information in huge quantities. We should also recognize that calls for government-induced content moderation will make these platforms battlegrounds for a perpetual intensifying conflict over who gets to silence whom. Finally, we need a renewed affirmation of Section 230 of the 1996 Telecommunications Act, which shields internet intermediaries from liability for users' speech. Contrary to Facebook's call for government-supervised content regulation, we need to keep platforms, not the state, responsible for finding the optimal balance between content moderation, freedom of expression, and economic value. The alternative of greater government regulation would absolve social media companies of market responsibility for their decisions and would probably lead them to exclude and suppress even more legal speech than they do now. It is the moral panic and proposals for regulation that threaten freedom and democracy.

“Social media make human interactions hypertransparent.”

## INTRODUCTION

In a few short years, social media platforms have gone from being shiny new paragons of the internet’s virtue to globally despised scourges. Once credited with fostering a global civil society and bringing down tyrannical governments, they are now blamed for an incredible assortment of social ills. In addition to legitimate concerns about data breaches and privacy, other ills—hate speech, addiction, mob violence, and the destruction of democracy itself—are all being laid at the doorstep of social media platforms.

Why are social media blamed for these ills? The human activities that are coordinated through social media, including negative things such as bullying, gossiping, rioting, and illicit liaisons, have always existed. In the past, these interactions were not as visible or accessible to society as a whole. As these activities are aggregated into large-scale, public commercial platforms, however, they become highly visible to the public and generate storable, searchable records. In other words, social media make human interactions hypertransparent.<sup>1</sup>

This new hypertransparency of social interaction has powerful effects on the dialogue about regulation of communications. It lends itself to the idea that social media causes the problems that it reveals and that society can be altered or engineered by meddling with the intermediaries who facilitate the targeted activities. Hypertransparency generates what I call the fallacy of displaced control. Society responds to aberrant behavior that is revealed through social media by demanding regulation of the intermediaries instead of identifying and punishing the individuals responsible for the bad acts. There is a tendency to go after the public manifestation of the problem on the internet, rather than punishing the undesired behavior itself. At its worst, this focus on the platform rather than the actor promotes the dangerous idea that government should regulate generic technological capabilities rather than bad behavior.

Concerns about foreign interference and

behavioral advertising brought a slowly simmering social media backlash to a boil after the 2016 election. As this reaction enters its third year, it is time to step back and offer some critical perspective and an assessment of where free expression fits into this picture. As hypertransparency brings to public attention disturbing, and sometimes offensive, content, a moral panic has ensued—one that could lead to damaging regulation and government oversight of private judgment and expression. Perhaps policy changes are warranted, but the regulations being fostered by the current social climate are unlikely to serve our deepest public values.

## MORAL PANIC

The assault on social media constitutes a textbook case of moral panic. Moral panics are defined by sociologists as “the outbreak of moral concern over a supposed threat from an agent of corruption that is out of proportion to its actual danger or potential harm.”<sup>2</sup> While the problems noted may be real, the claims “exaggerate the seriousness, extent, typicality and/or inevitability of harm.” In a moral panic, sociologist Stanley Cohen says, “the untypical is made typical.”<sup>3</sup> The exaggerations build upon themselves, amplifying the fears in a positive feedback loop. Purveyors of the panic distort factual evidence or even fabricate it to justify (over)reactions to the perceived threat. One of the most destructive aspects of moral panics is that they frequently direct outrage at a single easily identified target when the real problems have more complex roots. A sober review of the claims currently being advanced about social media finds that they tick off all these boxes.

## Fake News!

Social media platforms are accused of generating a cacophony of opinions and information that is degrading public discourse. A quote from a respected media scholar summarizes the oft-repeated view that social media platforms have an intrinsically negative impact on our information environment:

“In human history, what public medium has not mixed fact with fiction?”

An always-on, real-time information tsunami creates the perfect environment for the spread of falsehoods, conspiracy theories, rumors, and “leaks.” Unsubstantiated claims and narratives go viral while fact checking efforts struggle to keep up. Members of the public, including researchers and investigative journalists, may not have the expertise, tools, or time to verify claims. By the time they do, the falsehoods may have already embedded themselves in the collective consciousness. Meanwhile, fresh scandals or outlandish claims are continuously raining down on users, mixing fact with fiction.<sup>4</sup>

In this view, the serpent of social media has driven us out of an Eden of rationality and moderation. In response, one might ask: in human history, what public medium has *not* mixed fact with fiction, has *not* created new opportunities to spread falsehoods, or has *not* created new challenges for verification of fact? Similar accusations were levelled against the printing press, the daily newspaper, radio, and television; the claim that social media are degrading public discourse exaggerates both the uniqueness and the scope of the threat.

### Addiction and Extremism

A variant on this theme links the ad-driven business model of social media platforms to an inherently pathological distortion of the information environment: as one pundit wrote, “YouTube leads viewers down a rabbit hole of extremism, while Google racks up the ad sales.”<sup>5</sup> A facile blend of pop psychology and pop economics equates social media engagement to a dopamine shot for the user and increasing ad revenue for the platform. The way to prolong and promote such engagement, we are told, is to steer the user to increasingly extreme content. Any foray into the land of YouTube videos is a one-way ticket to beheadings, Alex Jones, flat-earthism, school-shooting denial, Pepe the Frog, and radical vegans. No more kittens, dog tricks, or baby pictures: for some unspecified

reason, those nice things are no longer what the platform delivers.

In the quote below, an academic evokes all the classical themes of media moral panics—addiction, threats to public health, and a lack of confidence in the agency of common people—into a single indictment of YouTube algorithmic recommendations:

Human beings have many natural tendencies that need to be vigilantly monitored in the context of modern life. For example, our craving for fat, salt and sugar, which served us well when food was scarce, can lead us astray in an environment in which fat, salt and sugar are all too plentiful and heavily marketed to us. So too our natural curiosity about the unknown can lead us astray on a website that leads us too much in the direction of lies, hoaxes and misinformation. In effect, YouTube has created a restaurant that serves us increasingly sugary, fatty foods, loading up our plates as soon as we are finished with the last meal.<sup>6</sup>

Another social media critic echoed similar claims:

Every pixel on every screen of every Internet app has been tuned to influence users’ behavior. Not every user can be influenced all the time, but nearly all users can be influenced some of the time. In the most extreme cases, users develop behavioral addictions that can lower their quality of life and that of family members, co-workers and close friends.<sup>7</sup>

If one investigates the “science” behind these claims, however, one finds little to differentiate social media addiction from earlier panics about internet addiction, television addiction, video game addiction, and the like. The evidence for the algorithmic slide toward media fat, salt, and sugar traces back to one man, Jonathan Albright of Columbia University’s Tow Center, and it is very difficult

“In the ultramoderated world that many of the social media critics seem to be advocating, important minority-viewpoint content is as likely to be targeted as terrorist propaganda and personal harassment.”

to find any published, peer-reviewed academic research from Albright. All one can find is a blog post on *Medium*, describing “the network of YouTube videos users are exposed to after searching for ‘crisis actor’ following the Parkland event.”<sup>8</sup> In other words, the blog reports the results of one search and one selected search phrase; there is no description of a methodology nor is there any systematic conceptualization or argumentation about the causal linkage between YouTube’s business model and the elevation of extreme and conspiratorial content. Yet Albright’s claims echoed through the *New York Times* and dozens of other online media outlets.

The psychological claims also seem to suffer from a moral panic bias. According to Courtney Seiter, a psychologist cited by some of the critics, the oxytocin and dopamine levels generated by social media use generate a positive “hormonal spike equivalent to [what] some people [get] on their wedding day.” She goes on to say that “all the goodwill that comes with oxytocin—lowered stress levels, feelings of love, trust, empathy, generosity—comes with social media, too . . . between dopamine and oxytocin, social networking not only comes with a lot of great feelings, it’s also really hard to stop wanting more of it.”<sup>9</sup> The methodological rigor and experimental evidence behind these claims seems to be thin, but even so, wasn’t social media supposed to be a tinderbox for hate speech? Somehow, citations of Seiter in attacks on social media seem to have left the trust, empathy, and generosity out of the picture.

The panic about elevating conspiratorial and marginalized content is especially fascinating. We are told in terms reminiscent of the censorship rationalizations of authoritarian governments that social media empowers the fringes and so threatens social stability. Yet for decades, mass media have been accused of appealing to the mainstream taste and of marginalizing anything outside of it. Indeed, in the 1970s, progressives tried to force media outlets to include marginalized voices in their channel lineup through public

access channels. Nowadays, apparently, the media system is dangerous because it does precisely the opposite.

But the overstatement of this claim should be evident. Major advertisers come down hard on the social platforms very quickly when their pitches are associated with crazies, haters, and blowhards, leading to algorithmic adjustments that suppress marginal voices. Users’ ability to “report” offensive content is another important form of feedback. But this has proven to cut both ways: lots of interesting but racy or challenging content gets suppressed. Some governments have learned how to game organized content moderation to yank messages exposing their evil deeds. (See the discussion of Facebook and Myanmar in the next section.) In the ultramoderated world that many of the social media critics seem to be advocating, important minority-viewpoint content is as likely to be targeted as terrorist propaganda and personal harassment.

**MURDER, HATE SPEECH, AND ETHNIC CLEANSING.** Another key exhibit in the case against social media pins the responsibility for ethnic cleansing in Myanmar, and similar incitement tragedies in the developing world, on Facebook. In this case, as in most of the other concerns, there is substance to the claim but its use and framing in the public discourse seems both biased and exaggerated. In Myanmar, the Facebook platform seems to have been systematically utilized as part of a state-sponsored campaign to target the Rohingya Muslim minority.<sup>10</sup> The government and its allies incited hatred against them, while censoring activists and journalists documenting state violence, by reporting their work as offensive content or in violation of community standards. At the same time, the government-sponsored misinformation and propaganda against the Rohingya managed to avoid the scrutiny applied to the expression of human-rights activists. Social media critics also charged that the Facebook News Feed’s tendency to promote already popular content allowed posts inciting violence against the minority to go viral. As a result, Facebook is

blamed for the tragedies in Myanmar. I have encountered people in the legal profession who would like to bring a human-rights lawsuit against Facebook.<sup>11</sup> If any criticism can be leveled at Facebook's handling of genocidal propaganda in Myanmar, it is that Facebook's moderation process is too deferential to governments. This, however, militates against greater state regulation.

But these claims show just how displaced the moral panic is. Why is so much attention being focused on Facebook and not on the crimes of a state actor? Yes, Myanmar military officers used Facebook (and other media) as part of an anti-Rohingya propaganda campaign. If the Burmese generals used telephones or text messages to spread their poison, are they going to blame those service providers or technologies? How about roads, which were undoubtedly used by the military to oppress Rohingya? In fact, violent conflict between Rohingya Muslims and Myanmar's majority population goes back to 1948, when the country achieved independence from the British and the new government denied citizenship to the Rohingya. A nationalist military coup in 1962 targeted them as a threat to the new government's concept of national identity; the army closed Rohingya social and political organizations, expropriated Rohingya businesses, and detained dissenters. It went on to regularly kill, torture, and rape Rohingya people.

Facebook disabled the accounts of the military propagandists once it understood the consequences of their misuse, although this happened much more slowly than critics would have liked. What's remarkable about the discussion of Facebook, however, is the way attention and responsibility for the oppression has been diverted away from a military dictatorship engaged in a state-sponsored campaign of ethnic cleansing, propaganda, and terror to a private foreign social media platform. In some cases, the discussion seems to imply that the absence of Facebook from Myanmar would solve, or even improve, the conflict that has been going on for 70 years. It is worth remembering that Facebook's status

as an *external* platform not under the control of the local government was the only thing that made it possible to intervene at all. Interestingly, the *New York Times* article that broke this story notes that pro-democracy officials in Myanmar say Facebook was essential for the democratic transition that brought them into office in 2015.<sup>12</sup> This claim is as important (and as unverified and possibly untestable) as the claim that it is responsible for ethnic cleansing. But it hasn't gotten any play lately.

**REVIVING THE RUSSIAN MENACE.** Russia-sponsored social media use during the 2016 election provides yet another example of the moral panic around social media and the avalanche of bitter exaggeration that goes with it. Indeed, the 2016 election marks the undisputed turning point in public attitudes toward social media. For many Americans, the election of Donald Trump came as a shocking and unpleasant surprise. In searching for an explanation of what initially seemed inexplicable, however, the nexus between the election results, Russian influence operations, and social media has become massively inflated. It has become too convenient to overlook Trump's complete capture of the Republican Party and his ability to capitalize on nationalistic and hateful themes that conservative Republicans had been cultivating for decades. The focus on social media continues to divert our attention from the well-understood negatives of Hillary Clinton as well as the documented impact of James Comey's decision to reopen the FBI investigation of Clinton's emails at a critical period in the presidential campaign. It overlooks, too, the strength of the Bernie Sanders challenge and the way the Clinton-controlled Democratic National Committee alienated his supporters. It also tends to downplay the linkages that existed between Trump's campaign staff, advisers, and Russia that had nothing to do with social media influence.

How much more comforting it was to focus on a foreign power and its use of social media than to face up to the realities of a politically polarized America and the way politicians

“Pro-democracy officials in Myanmar say Facebook was essential for the democratic transition that brought them into office in 2015.”

“These messages are attributed enormous power, as if they are the only ones anyone sees.”

and their crews peddle influence to a variety of foreign states and interests.<sup>13</sup> As this displacement of blame developed, references to Russian *information operations* uniformly became references to Russian *interference* in the elections.<sup>14</sup> Interference is a strong word—it makes it seem as if leaks of real emails and a disinformation campaign of Twitter bots and Facebook accounts were the equivalent of stuffing ballot boxes, erasing votes, hacking election machines, or forcibly blocking people from the polls. As references to foreign election interference became deeply embedded in the public discourse, the threat could be further inflated to one of national security. And so suddenly, the regulation of political speech got on the agenda of Congress, and millions of liberals and progressives became born-again Cold Warriors, all too willing to embrace nationalistic controls on information flows.

In April 2016 hackers employed by the Russian government compromised several servers belonging to the Democratic National Committee, exfiltrated a trove of internal communications, and published them via Wikileaks using a “Guccifer 2.0” alias.<sup>15</sup> The emails leaked by the Russians were not made up by the Russians; they were real. What if they had been leaked by a 21st-century Daniel Ellsberg instead of the Russians? Would that also be considered election interference? Disclosures of compromising information (e.g., Trump’s *Access Hollywood* tape) have a long history in American politics. Is that election interference? How much of the cut-and-thrust of an open society’s media system, and how many whistleblowers, are we willing to muzzle in this moral panic?

**THE DEATH OF DEMOCRACY.** Some critics go so far as to claim that democracy itself is threatened by the existence of open social media platforms. “[Facebook] has swallowed up the free press, become an unstoppable private spying operation and undermined democracy. Is it too late to stop it?” asks the subtitle of one typical article.<sup>16</sup> This critique is as common as it is inchoate. In its worst and most simple-minded form, the mere ability of foreign governments to put messages

on social media platforms is taken as proof that the entire country is being controlled by them. These messages are attributed enormous power, as if they are the only ones anyone sees; as if foreign governments don’t routinely buy newspaper ads, hire Washington lobbyists, or fund nonprofits and university programs. Worse still, those of this mindset equate messages with weapons in ceaseless “information warfare.” It is claimed that social media are being, or have been, “weaponized”—a transitive verb that was popularized after being applied to the 9/11 attackers’ use of civilian aircraft to murder thousands of people.<sup>17</sup> Users of this term show not the slightest embarrassment at a possible overstatement implicit in the comparison.

Cybersecurity writer Thomas Rid made the astounding assertion that the most “open and liberal social media platform” (Twitter) is “a threat to open and liberal democracy” precisely because it is open and liberal, thus implying that free expression is a national security threat.<sup>18</sup> In a *Time Magazine* cover story, a former Facebook executive complained that Facebook has “aggravated the flaws in our democracy while leaving citizens ever less capable of thinking for themselves.”<sup>19</sup> The nature of this threat is never scientifically documented in terms of its actual effect on voting patterns or political institutions. The only evidence offered is simple counts of the number of Russian trolls and bots and their impressions—numbers that look unimpressive compared to the spread of a single Donald Trump tweet. What we don’t often hear is that social media is the most important source of news for only 14 percent of the population. Research by two economists concluded that “. . . social media have become an important but not dominant source of political news and information. Television remains more important by a large margin.” They also conclude that there is no statistically significant correlation between social media use and those who draw ideologically aligned conclusions from their exposure to news.<sup>20</sup>

The most disturbing element of the “threat to democracy” argument is the way it

militarizes public discourse. The view of social media as information warfare seems to go hand-in-hand with the contradictory idea that imposing more regulation by the nation-state will “disarm” information and parry this threat to democracy. In advancing what they think of as sophisticated claims that social media are being weaponized, the joke is on our putative cybersecurity experts: it is Russian and Chinese doctrine that the free flow of information across borders is a subversive force that challenges their national sovereignty. This doctrine, articulated in a code of conduct by the Shanghai Cooperation Organization, was designed to rationalize national blocking and filtering of internet content.<sup>21</sup> By equating the influence that occurs via exchanges of ideas, information, and propaganda with war and violence, these pundits pose a more salient danger to democracy and free speech than any social media platform.

Any one of these accusations—the destruction of public discourse, responsibility for ethnic cleansing and hate speech, abetting a Russian national security threat, and the destruction of democracy—would be serious enough. Their combination in a regularly repeated catechism constitutes a moral panic. Moral panics should inspire caution because they produce policy reactions that overshoot the mark. A fearful public can be stampeded into legal or regulatory measures that serve a hidden agenda. Targeted actors can be scapegoated and their rights and interests discounted. Freedom-enhancing policies and proportionate responses to problems never emerge from moral panics.

### Media Panics in the Past

One antidote to moral panic is historical perspective. Media studies professor Kirsten Drotner wrote, “[E]very time a new mass medium has entered the social scene, it has spurred public debates on social and cultural norms, debates that serve to reflect, negotiate and possibly revise these very norms . . . In some cases, debate of a new medium brings about—indeed changes into—heated,

emotional reactions . . . what may be defined as a media panic.”<sup>22</sup> We need to understand that we are in the midst of one of these renegotiations of the norms of public discourse and that the process has tipped over into media panic—one that demonizes social media generically.

We can all agree that literacy is a good thing. In the 17th and 18th centuries, however, some people considered literacy’s spread subversive or corrupting. The expansion of literacy from a tiny elite to the general population scared a lot of conservatives. It meant not only that more people could read the Bible, but also that they could read radical liberal tracts such as Thomas Paine’s *Rights of Man*. Those who feared wider literacy believed that it generated conflict and disruption. In fact, it already had. The disintermediation of authority over the interpretation of the written word by the printing press and by wider literacy created centrifugal forces. Protestants had split with Catholics, and later, different Protestant sects formed around different interpretations of scripture. Later, in the 17th and 18th centuries, the upper class and the religious also complained about sensationalistic printed broadsheets and printed ballads that appealed to the “baser instincts” of the public. Commercial media that responded to what the people wanted were not perceived kindly by those who thought they knew best. Yet are these observations an argument for keeping people illiterate? If not, then what, exactly, do these concerns militate for? A controlled, censored press? A press licensed in “the public interest”? Who in those days would have been made the arbiter of public interest? The Pope? Absolutist kings?

Radio broadcasting was an important revolution in mass media technology. It seems to have escaped the intense, concentrated panic we are seeing around contemporary social media, but in the United States, where broadcasting had relatively free and commercial origins, those in power felt threatened by its potential to evolve into an independent medium. Thomas Hazlett has documented the way the 1927 Federal Radio Act and the regulatory

“Moral panics should inspire caution because they produce policy reactions that overshoot the mark.”

“The retailers and instigators of media panic generally pose the biggest threat to free expression and democracy.”

commission it created (later to become the Federal Communications Commission) nationalized the airwaves in order to keep the new medium licensed and under the thumb of Congress.<sup>23</sup> Numerous scholarly accounts have shown how the public-interest licensing regime erected after the federal takeover of the airwaves led to a systematic exclusion of diverse voices, from socialists to African Americans to labor unions.<sup>24</sup>

There is another relevant parallel between radio and social media. Totalitarian dictatorships, particularly Nazi Germany, employed radio broadcasting extensively in the 1930s. Those uses, some of which sparked the birth of modern communications effects research, were much scarier than the uses of social media by today's dictatorships and illiberal democracies. But oddly, our current panic tends to promote and support precisely the types of regulation and control favored by those very same modern dictatorships and illiberal democracies: centralized content moderation and blocking by the state and holding social media platforms responsible for the postings of their users.

Comic books generated a media panic in the 1940s and 50s.<sup>25</sup> A critic of American commercial culture, Frederic Wertham, believed that comic books encouraged juvenile delinquency and subverted the morality of children for the sake of profit. The presence of weirdness, violence, horror, and sexually tinged images led to charges that the comics were dangerous, addictive, and catered to baser instincts. A comic-book scare ensued, complete with a flood of newspaper stories, Congressional hearings, and a transformation of the comic book industry. The comic-book scare seems to have pioneered the three themes that characterize so much public discourse around new media in the 20th century: anti-commercialism, protecting children, and addiction. All are echoed in the current fight over social media. The same themes sounded in policy battles over television. Television's status as a cause of violence was debated and researched endlessly. Its pollution of public discourse, the way it “cultivated” inaccurate and harmful stereotypes, and

its addictive qualities were constant sources of discussion.<sup>26</sup> Again the similarity to current debates about social media is apparent.

In examining historical cases, it becomes apparent that it is the retailers and instigators of media panic who generally pose the biggest threat to free expression and democracy. For at their root, attacks on new media, past and present, are expressions of fear: fear of empowering diverse and dissonant voices, the elites' fears over losing hegemony over public discourse, and a lack of confidence in the ability of ordinary people to control their “baser instincts” or make sense of competing claims. The more sophisticated variants of these critiques are rationalizations of paternalism and authoritarianism. In the social media panic, we have both conservative *and* liberal elites recoiling from the prospect of a public sphere over which they have lost control, and both are preparing the way for regulatory mechanisms that can tame diversity, homogenize output, and maintain their established place in society.

### What's Broken?

A recent exchange on Twitter exposed the policy vacuity of those leading the social media moral panic. Kara Swisher, a well-known tech journalist with more than a million followers, tweeted to Jack Dorsey, the CEO of Twitter:

Overall here is my mood and I think a lot of people when it comes to fixing what is broke about social media and tech: Why aren't you moving faster? Why aren't you moving faster? Why aren't you moving faster?<sup>27</sup>

Swisher's impatient demand for fast action seemed to assume that the solutions to social media's ills were obvious. I tweeted in reply, asking what “fix” she wanted to implement so quickly. There was no answer.

Here is the diagnosis I would offer. What is “broken” about social media is exactly the same thing that makes it useful, attractive, and commercially successful: it is incredibly effective at facilitating discoveries and exchanges



of information among interested parties at unprecedented scale. As a direct result of that, there are more informational interactions than ever before and more mutual exchanges between people. This human activity, in all its glory, gore, and squalor, generates storable, searchable records, and its users leave attributable tracks everywhere. As noted before, the emerging new world of social media is marked by hypertransparency.

From the standpoint of free expression and free markets there is nothing inherently broken about this; on the contrary, most of the critics are unhappy precisely *because* the model is working: it is unleashing all kinds of expression and exchanges, and making tons of money at it to boot. But two distinct sociopolitical pathologies are generated by this. The first is that, by exposing all kinds of deplorable uses and users, it tends to funnel outrage at these manifestations of social deviance toward the platform providers. A man discovers pedophiles commenting on YouTube videos of children and is sputtering with rage at . . . YouTube.<sup>28</sup> The second pathology is the idea that the objectionable behaviors can be engineered out of existence or that society as a whole can be engineered into a state of virtue by encouraging intermediaries to adopt stricter surveillance and regulation. Instead of trying to stop or control the objectionable behavior, we strive to control the communications intermediary that was used by the bad actor. Instead of eliminating the crime, we propose to deputize the intermediary to recognize symbols of the crime and erase them from view. It's as though we assume that life is a screen, and if we remove unwanted things from our screens by controlling internet intermediaries, then we have solved life's problems. (And even as we do this, we hypocritically complain about China and its alleged development of an all-embracing social credit system based on online interactions.)

The reaction against social media is thus based on a false premise and a false promise. The false premise is that the creators of tools that enable public interaction at scale are primarily *responsible* for the existence of the

behaviors and messages so revealed. The false promise is that by pushing the platform providers to block content, eliminate accounts, or otherwise attack manifestations of social problems on their platforms, we are solving or reducing those problems. Combing these misapprehensions, we've tried to curb "new" problems by hiding them from public view.

The major platforms have contributed to this pathology by taking on ever-more-extensive content-moderation duties. Because of the intense political pressure they are under, the dominant platforms are rapidly accepting the idea that they have overarching social responsibilities to shape user morals and shape public discourse in politically acceptable ways. Inevitably, due to the scale of social media interactions, this means increasingly automated or algorithmic forms of regulation, with all of its rigidities, stupidities, and errors. But it also means massive investments in labor-intensive manual forms of moderation.<sup>29</sup>

The policy debate on this topic is complicated by the fact that internet intermediaries cannot really avoid taking on some optional content regulation responsibilities beyond complying with various laws. Their status as multisided markets that match providers and seekers of information requires it.<sup>30</sup> Recommendations based on machine learning guide users through the vast, otherwise intractable amount of material available. These filters vastly improve the value of a platform to a user, but they also indirectly shape what people see, read, and hear. They can also, as part of their attempts to attract users and enhance the platforms' value to advertisers, discourage or suppress messages and forms of behavior that make their platforms unpleasant or harmful places. This form of content moderation is outside the scope of the First Amendment's legal protections because it is executed by a private actor and falls within the scope of editorial discretion.

### What's the Fix?

Section 230 of the Communications Decency Act squared this circle by immunizing

“The reaction against social media is thus based on a false premise and a false promise.”

“What stance should advocates of both free expression and free markets take with respect to social media?”

information service providers who did nothing to restrict or censor the communications of the parties using their platforms (the classical “neutral conduit” or common-carrier concept), while also immunizing information service providers who assumed some editorial responsibilities (e.g., to restrict pornography and other forms of undesirable content). Intermediaries who did nothing were (supposed to be) immunized in ways that promoted freedom of expression and diversity online; intermediaries who were more active in managing user-generated content were immunized to enhance their ability to delete or otherwise monitor “bad” content without being classified as publishers and thus assuming responsibility for the content they did not restrict.<sup>31</sup>

It is clear that this legal balancing act, which worked so well to make the modern social media platform successful, is breaking down. Section 230 is a victim of its own success. Platforms have become big and successful in part because of their Section 230 freedoms, but as a result they are subject to political and normative pressures that confer upon them de facto responsibility for what their users read, see, and do. The threat of government intervention is either lurking in the background or being realized in certain jurisdictions. Fueled by hypertransparency, political and normative pressures are making the pure, neutral, non-discriminatory platform a thing of the past.

The most common proposals for fixing social media platforms all seem to ask the platforms to engage in *more* content moderation and to ferret out unacceptable forms of expression or behavior. The political demand for more-aggressive content moderation comes primarily from a wide variety of groups seeking to suppress specific kinds of content that is objectionable to them. Those who want less control or more toleration suffer from the diffuse costs/concentrated benefit problem familiar to us from the economic analysis of special interest groups: that is, toleration benefits everyone a little and its presence is barely noticeable until it is lost; suppression, on the other hand, offers powerful and immediate

satisfaction to a few highly motivated actors.<sup>32</sup>

At best, reformers propose to rationalize content moderation in ways designed to make its standards clearer, make their application more consistent, and make an appeals process possible.<sup>33</sup> Yet this is unlikely to work unless platforms get the backbone to strongly assert their rights to set the criteria, stick to them, and stop constantly adjusting them based on the vagaries of daily political pressures. At worst, advocates of more content moderation are motivated by a belief that greater content control will reflect their own personal values and priorities. But since calls for tougher or more extensive content moderation come from all ideological and cultural directions, this expectation is unrealistic. It will only lead to a distributed form of the heckler’s veto, and a complete absence of predictable, relatively objective standards. It is not uncommon for outrage at social media to lead in contradictory directions. A reporter for *The Guardian*, for example, is outraged that Facebook has an ad-targeting category for “vaccine controversies” and flogs the company for allowing anti-vaccination advocates to form closed groups that can reinforce those members’ resistance to mainstream medical care.<sup>34</sup> However, there is no way for Facebook to intervene without profiling their users as part of a specific political movement deemed to be wrong, and then suppressing their communications and their ability to associate based on that data. So, at the same time Facebook is widely attacked for privacy violations, it is also being asked to leverage its private user data to flag political and social beliefs that are deemed aberrant and to suppress users’ ability to associate, connect with advertisers, or communicate among themselves. In this combination of surveillance and suppression, what could possibly go wrong?

What stance should advocates of both free expression and free markets take with respect to social media?

First, there needs to be a clearer articulation of the tremendous value of platforms based on their ability to match seekers and providers of information. There also needs to be explicit

advocacy for greater tolerance of the jarring diversity revealed by these processes. True liberals need to make it clear that social media platforms cannot be expected to bear the main responsibility for sheltering us from ideas, people, messages, and cultures that we consider wrong or that offend us. Most of the responsibility for what we see and what we avoid should lie with us. If we are outraged by seeing things we don't like in online communities comprised of billions of people, we need to stop misdirecting that outrage against the platforms that happen to expose us to it. Likewise, if the exposed behavior is illegal, we need to focus on identifying the perpetrators and holding them accountable. As a corollary of this attitudinal change, we also need to show that the hypertransparency fostered by social media can have great social value. As a simple example of this, research has shown that the much-maligned rise of platforms matching female sex workers with clients is statistically correlated with a decrease in violence against women—precisely because it took sex work off the street and made transactions more visible and controllable.<sup>35</sup>

Second, free-expression supporters need to actively challenge those who want content moderation to go further. We need to expose the fact that they are using social media as a means of reforming and reshaping society, wielding it like a hammer against norms and values they want to be eradicated from the world. These viewpoints are leading us down an authoritarian blind alley. They may very well succeed in suppressing and crippling the freedom of digital media, but they will not, and cannot, succeed in improving society. Instead, they will make social media platforms battlegrounds for a perpetual intensifying conflict over who gets to silence whom. This is already abundantly clear from the cries of discrimination and bias as the platforms ratchet up content moderation: the cries come from both the left and the right in response to moderation that is often experienced as arbitrary.

Finally, we need to mount a renewed and reinvigorated defense of Section 230. The case for Section 230 is simple: no alternative

promises to be intrinsically better than what we have now, and most alternatives are likely to be worse. The exaggerations generated by the moral panic have obscured the simple fact that moderating content on a global platform with billions of users is an extraordinarily difficult and demanding task. Users, not platforms, are the source of messages, videos, and images that people find objectionable, so calls for regulation ignore the fact that regulations don't govern a single supplier, but must govern millions, and maybe billions, of users. The task of flagging user-generated content, considering it, and deciding what to do about it is difficult and expensive. And is best left to the platforms.

However, regulation seems to be coming. Facebook CEO Mark Zuckerberg has published a blog post calling for regulating the internet, and the UK government has released a white paper, "Online Harms," that proposes the imposition of systematic liability for user-generated content on all internet intermediaries (including hosting companies and internet service providers).<sup>36</sup>

At best, a system of content regulation influenced by government is going to look very much like what is happening now. Government-mandated standards for content moderation would inevitably put most of the responsibility for censorship on the platforms themselves. Even in China, with its army of censors, the operationalization of censorship relies heavily on the platform operators. In the tsunami of content unleashed by social media, prior restraint by the state is not really an option. Germany responded in a similar fashion with the 2017 *Netzwerkdurchsetzungsgesetz*, or Network Enforcement Act (popularly known as NetzDG or the Facebook Act), a law aimed at combating agitation, hate speech, and fake news in social networks.

The NetzDG law immediately resulted in suppression of various forms of politically controversial online speech. Joachim Steinhöfel, a German lawyer concerned by Facebook's essentially jurisprudential role under NetzDG, created a "wall of shame" containing legal content suppressed by NetzDG.<sup>37</sup> Ironically,

“The NetzDG law immediately resulted in suppression of various forms of politically controversial online speech.”

“Governmental involvement in content regulation would have to conform to the First Amendment.”

German right-wing nationalists who suffered takedowns under the new law turned the law to their advantage by using it to suppress critical or demeaning comments about themselves. “Germany’s attempt to regulate speech online has seemingly amplified the voices it was trying to diminish,” claims an article in *The Atlantic*.<sup>38</sup> As a result of one right-wing politician’s petition, Facebook must ensure that individuals in Germany cannot use a VPN to access illegal content. Yet still, a report by an anti-hate-speech group that supports the law argues that it has been ineffective. “There have been no fines imposed on companies and little change in overall takedown rates.”<sup>39</sup>

Abandoning intermediary immunities would make the platforms even more conservative and more prone to disable accounts or take down content than they are now. In terms of costs and legal risks, it will make sense for them to err on the safe side. When intermediaries are given legal responsibility, conflicts about arbitrariness and false positives don’t go away, they intensify. In authoritarian countries, platforms will be merely be indirect implementers of national censorship standards and laws.

On the other hand, U.S. politicians face a unique and interesting dilemma. If they think they can capitalize on social media’s travails with calls for regulation, they must understand that governmental involvement in content regulation would have to conform to the First Amendment. This would mean that all kinds of content that many users don’t want to see, ranging from hate speech to various levels of nudity, could no longer be restricted because they are *not* strictly illegal. Any government interventions that took down postings or deleted accounts could be litigated based on a First Amendment standard. Ironically, then, a governmental takeover of content regulation responsibilities in the United States would have to be far more liberal than the status quo. Avoidance of this outcome was precisely why Section 230 was passed in the first place.

From a pure free-expression standpoint, a First Amendment approach would be a good thing. But from a free-association and

free-market standpoint, it would not. Such a policy would literally force all social media users to be exposed to things they didn’t want to be exposed to. It would undermine the economic value of platforms by decapitating their ability to manage their matching algorithms, shape their environment, and optimize the tradeoffs of a multisided market. Given the current hue and cry about all the bad things people are seeing and doing on social media, a legally driven, permissive First Amendment standard does not seem like it would make anyone happy.

Advocates of expressive freedom, therefore, need to reassert the importance of Section 230. Platforms, not the state, should be responsible for finding the optimal balance between content moderation, freedom of expression, and the economic value of platforms. The alternative of greater government regulation would absolve the platforms of market responsibility for their decisions. It would eliminate competition among platforms for appropriate moderation standards and practices and would probably lead them to exclude and suppress even more legal speech than they do now.

## CONCLUSION

Content regulation is only the most prominent of the issues faced by social media platforms today; they are also implicated in privacy and competition-policy controversies. But social media content regulation has been the exclusive focus of this analysis. Hypertransparency and the subsequent demand for content control it creates are the key drivers of the new media moral panic. The panic is feeding upon itself, creating conditions for policy reactions that overlook or openly challenge values regarding free expression and free enterprise. While there is a lot to dislike about Facebook and other social media platforms, it’s time we realized that a great deal of that negative reaction stems from an information society contemplating manifestations of itself. It is not an exaggeration to say that we are blaming the mirror for what we see in it. Section 230 is still surprisingly relevant to this dilemma. As

a policy, Section 230 was not a form of infant industry protection that we can dispense with now, nor was it a product of a utopian inebriation with the potential of the internet. It was a very clever way of distributing responsibility for content governance in social media. If we stick with this arrangement, learn more tolerance, and take more responsibility for what we see and do on social media, we can respond to the problems while retaining the benefits.

## NOTES

1. Milton L. Mueller, "Hyper-transparency and Social Control: Social Media as Magnets for Regulation," *Telecommunications Policy* 39, no. 9 (2015): 804–10.
2. Erich Goode and Nachman Ben-Yehuda, "Grounding and Defending the Sociology of Moral Panic," chap. 2 in *Moral Panic and the Politics of Anxiety*, ed. Sean Patrick Hier (Abingdon: Routledge, 2011).
3. Stanley Cohen, *Folk Devils and Moral Panics* (Abingdon: Routledge, 2011).
4. Ronald J. Deibert, "The Road to Digital Unfreedom: Three Painful Truths about Social Media," *Journal of Democracy* 30, no. 1 (2019): 25–39.
5. Zeynep Tufekci, "YouTube, the Great Radicalizer," *New York Times*, March 10, 2018.
6. Tufekci, "YouTube, the Great Radicalizer."
7. Roger McNamee, "I Mentored Mark Zuckerberg. I Loved Facebook. But I Can't Stay Silent about What's Happening," *Time Magazine*, January 17, 2019.
8. Jonathan Albright, "Untrue-Tube: Monetizing Misery and Disinformation," *Medium*, February 25, 2018.
9. Courtney Seiter, "The Psychology of Social Media: Why We Like, Comment, and Share Online," *Buffer*, August 20, 2017.
10. Paul Mozur, "A Genocide Incited on Facebook, With Posts from Myanmar's Military," *New York Times*, October 15, 2018.
11. Ingrid Burrington, "Could Facebook Be Tried for Human-Rights Abuses?," *The Atlantic*, December 20, 2017.
12. Burrington, "Could Facebook Be Tried for Human-Rights Abuses?"
13. For a discussion of Michael Flynn's lobbying campaign for the Turkish government and Paul Manafort's business in Ukraine and Russia, see Rebecca Kheel, "Turkey and Michael Flynn: Five Things to Know," *The Hill*, December 17, 2018; and Franklin Foer, "Paul Manafort, American Hustler," *The Atlantic*, March 2018.
14. See, for example, "Minority Views to the Majority-produced 'Report on Russian Active Measures, March 22, 2018'" of the Democratic representatives from the United States House Permanent Select Committee on Intelligence (USHPSCI), March 26, 2018.
15. Indictment at 11, *U.S. v. Viktor Borisovich Netyksbo et al.*, Case 1:18-cr-00032-DLF (D.D.C. filed Feb. 16, 2018).
16. Matt Taibbi, "Can We Be Saved from Facebook?," *Rolling Stone*, April 3, 2018.
17. Peter W. Singer and Emerson T. Brooking, *Like War: The Weaponization of Social Media* (New York: Houghton Mifflin Harcourt, 2018).
18. Thomas Rid, "Why Twitter Is the Best Social Media Platform for Disinformation," *Motherboard*, November 1, 2017.
19. McNamee, "I Mentored Mark Zuckerberg. I Loved Facebook. But I Can't Stay Silent about What's Happening."
20. Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 31, no. 2 (2017): 211–36.
21. Sarah McKune, "An Analysis of the International Code of Conduct for Information Security," CitizenLab, September 28, 2015.
22. Kirsten Drotner, "Dangerous Media? Panic Discourses and Dilemmas of Modernity," *Paedagogica Historica* 35, no. 3 (1999): 593–619.
23. Thomas W. Hazlett, "The Rationality of US Regulation of

- the Broadcast Spectrum,” *Journal of Law and Economics* 33, no. 1 (1990): 133–75.
24. Robert McChesney, *Telecommunications, Mass Media and Democracy: The Battle for Control of U.S. Broadcasting, 1928–1935* (New York: Oxford, 1995).
25. Fredric Wertham, *Seduction of the Innocent* (New York: Rinehart, 1954); and David Hajdu, *The Ten-cent Plague: The Great Comic-book Scare and How It Changed America* (New York: Picador, 2009), <https://us.macmillan.com/books/9780312428235>.
26. “Like drug dealers on the corner, [TV broadcasters] control the life of the neighborhood, the home and, increasingly, the lives of children in their custody,” claimed a former FCC commissioner. Minow & LeMay, 1995. <http://www.washingtonpost.com/wp-srv/style/longterm/books/chap1/abandonedinthewasteland.htm>. Newton N. Minow & Craig L. LaMay, *Abandoned in the Wasteland* (New York: Hill and Wang, 1996).
27. Kara Swisher (@karaswisher), “Overall here is my mood and I think a lot of people when it comes to fixing what is broke about social media and tech: Why aren’t you moving faster? Why aren’t you moving faster? Why aren’t you moving faster?” Twitter post, February 12, 2019, 2:03 p.m., <https://twitter.com/karaswisher/status/1095443416148787202>.
28. Matt Watson, “Youtube Is Facilitating the Sexual Exploitation of Children, and It’s Being Monetized,” YouTube video, 20:47, “Matt’s What It Is,” February 27, 2019, <https://www.youtube.com/watch?v=O13G5A5w5Po>.
29. Casey Newton, “The Trauma Floor: The Secret Lives of Facebook Moderators in America,” *The Verge*, February 25, 2019.
30. Geoff Parker, Marshall van Alstyne, and Sangeet Choudhary, *Platform Revolution* (New York: W. W. Norton, 2016).
31. The Court in *Zeran v. America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997), said Sec. 230 was passed to “remove the disincentives to self-regulation created by the *Stratton Oakmont* decision.” In *Stratton Oakmont, Inc. v. Prodigy Services Co.*, (N.Y. Sup. Ct. 1995), a bulletin-board provider was held responsible for defamatory remarks by one of its customers because it made efforts to edit some of the posted content.
32. Robert D Tollison, “Rent Seeking: A Survey,” *Kyklos* 35, no. 4 (1982): 575–602.
33. See, for example, the “Santa Clara Principles on Transparency and Accountability in Content Moderation,” May 8, 2018, <https://santaclaraprinciples.org/>.
34. Julia Carrie Wong, “Revealed: Facebook Enables Ads to Target Users Interested in ‘Vaccine Controversies,’” *The Guardian* (London), February 15, 2019.
35. See Scott Cunningham, Gregory DeAngelo, and John Tripp, “Craigslist’s Effect on Violence against Women,” <http://scunning.com/craigslist110.pdf> (2017). See also Emily Witt, “After the Closure of Backpage, Increasingly Vulnerable Sex Workers Are Demanding Their Rights,” *New Yorker*, June 8, 2018.
36. Mark Zuckerberg, “Four Ideas to Regulate the Internet,” March 30, 2019; and UK Home Office, Department for Digital, Culture, Media & Sport, *Online Harms White Paper*, The Rt Hon. Sajid Javid MP, The Rt Hon. Jeremy Wright MP, April 8, 2019.
37. Joachim Nikolaus Steinhöfel, “Blocks & Hate Speech—Insane Censorship & Arbitrariness from FB,” Facebook Block - Wall of Shame, <https://facebook-sperre.steinhoefel.de/>.
38. Linda Kinstler, “Germany’s Attempt to Fix Facebook Is Backfiring,” *The Atlantic*, May 18, 2018.
39. William Echikson and Olivia Knodt, “Germany’s NetzDG: A Key Test for Combatting Online Hate,” CEPS Research Report no. 2018/09, November 2018.

## **RELATED PUBLICATIONS FROM THE CATO INSTITUTE**

**Why the Government Should Not Regulate Content Moderation of Social Media** by John Samples, Cato Institute Policy Analysis no. 865 (April 9, 2019)

**What to Do about the Emerging Threat of Censorship Creep on the Internet** by Danielle Keats Citron, Cato Institute Policy Analysis no. 828 (November 28, 2017)

**Commercial Speech and the Values of Free Expression** by Martin H. Redish, Cato Institute Policy Analysis no. 813 (June 19, 2017)

**Surveillance Takes Wing: Privacy in the Age of Police Drones** by Matthew Feeney, Cato Institute Policy Analysis no. 807 (December 13, 2016)

**Hate Speech Laws: Ratifying the Assassin's Veto** by Robert Corn-Revere, Cato Institute Policy Analysis no. 791 (May 24, 2016)

**Leashing the Surveillance State: How to Reform Patriot Act Surveillance Authorities** by Julian Sanchez, Cato Institute Policy Analysis no. 675 (May 16, 2011)

**Fairness 2.0: Media Content Regulation in the 21st Century** by Robert Corn-Revere, Cato Institute Policy Analysis no. 651 (November 10, 2009)

**Broadcast Localism and the Lessons of the Fairness Doctrine** by John Samples, Cato Institute Policy Analysis no. 639 (May 27, 2009)

**Caught in the Seamless Web: Does the Internet's Global Reach Justify Less Freedom of Speech?** by Robert Corn-Revere, Cato Institute Briefing Paper no. 71 (July 24, 2002)

**RECENT STUDIES IN THE  
CATO INSTITUTE POLICY ANALYSIS SERIES**

- 875. **The Community Reinvestment Act in the Age of Fintech and Bank Competition** by Diego Zuluaga (July 10, 2019)
- 874. **Closing Pandora’s Box: The Growing Abuse of the National Security Rationale for Restricting Trade** by Simon Lester and Huan Zhu (June 25, 2019)
- 873. **Immigration Wait Times from Quotas Have Doubled: Green Card Backlogs Are Long, Growing, and Inequitable** by David J. Bier (June 18, 2019)
- 872. **Is This Time Different? Schumpeter, the Tech Giants, and Monopoly Fatalism** by Ryan Bourne (June 17, 2019)
- 871. **Unplugging the Third Rail: Choices for Affordable Medicare** by John F. Early (June 6, 2019)
- 870. **Principles for the 2020 Surface Transportation Reauthorization** by Randal O’Toole (June 4, 2019)
- 869. **A Reform Agenda for the Next Indian Government** by Swaminathan S. Anklesaria Aiyar (May 21, 2019)
- 868. **Restoring Responsible Government by Cutting Federal Aid to the States** by Chris Edwards (May 20, 2019)
- 867. **Making Sense of the Minimum Wage: A Roadmap for Navigating Recent Research** by Jeffrey Clemens (May 14, 2019)
- 866. **Terrorists by Immigration Status and Nationality: A Risk Analysis, 1975–2017** by Alex Nowrasteh (May 7, 2019)
- 865. **Why the Government Should Not Regulate Content Moderation of Social Media** by John Samples (April 9, 2019)
- 864. **Overdosing on Regulation: How Government Caused the Opioid Epidemic** by Jeffrey Miron, Greg Sollenberger, and Laura Nicolae (February 14, 2019)
- 863. **How “Market Failure” Arguments Lead to Misguided Policy** by Ryan Bourne (January 22, 2019)