PUBLIC CALLS FOR CENSORSHIP AS BAD SPEECH

J.P. Messina[*]

Responsible speakers avoid trafficking in *bad speech*, that is, speech that they have reason to believe causes or constitutes net harm. Moreover, third parties have *prima facie* reason to suppress such speech. As recent events have made salient just how harmful speech can be, there has been a corresponding increase in calls to suppress or censor such speech. This article argues that there are three mechanisms by which calls to suppress bad speech themselves tend to cause or constitute harm. Paradoxically, then, those most concerned about the pernicious effects of bad speech ought to be especially reluctant to call for its suppression.

INTRODUCTION

Recent political events have tested the strength of the Western commitment to free speech principles. A tumultuous 2020 U.S. election (culminating in the Capitol riots of 2021), the ongoing COVID-19 pandemic, and Russia's war of aggression against Ukraine have likely altered American attitudes toward an ideal that has enjoyed broad bipartisan support since the McCarthy era. As support for free expression wanes or takes a different shape, social media companies have been under intense public and state pressure to more strictly moderate the content they host, and

[*] Assistant Professor of Philosophy, Purdue University (jpmessin@purdue.edu).

they have made extensive efforts to comply.[1] As protests over racial justice turned (perhaps justifiably) violent, panic over Critical Race Theory (CRT) has reached a fever pitch, bringing with it regular calls for schools or legislatures to remove books from public school libraries and curricula, subjecting secondary school curricula to greater legislative oversight.[2] And, in spite of a previously robust commitment to academic freedom, academic presses and journals have faced new pressure from scholars to refrain from publishing material thought harmful to marginalized groups and inimical to collective aims.

Naturally, Americans disagree about which speech now warrants suppression. But new survey data suggest that most think that at least some speech ought to be suppressed, be it by the state or by private parties. Here are a few examples of what Democrats and Republicans view as illegitimate exercises of their First Amendment rights, all drawn from the recent Knight-Ipsos survey on "Free Expression in America Post-2020":[3]

- kneeling or turning away during the national anthem (65% of Republicans, 43% overall);
- spreading misinformation about the 2020 election results online (80% of Democrats, 66% overall);
- taking part in racial justice protests over the summer of 2020 (44% of Republicans, 27% overall);
- spreading misinformation about the COVID-19 vaccine online (80% of Democrats, 70% overall).

It is important to avoid giving the sense that 2020 marks the end of a golden age of American tolerance for an age in which polarization and partisan animus are the norm. Indeed, another survey suggests that polarization around tolerable speech was already a main fixture of American civic life in 2017, before these precipitating events. For example, as of 2017:[4]

---

[1] *See* Rebecca Klar, *Feds Step up Pressure on Social Media over False COVID-19 Claims*, THE HILL (July 18, 2021, 11:00 AM), https://perma.cc/8TYL-CTYH.

[2] *See* Jeffrey Sachs, *Scope and Speed of Educational Gag Orders Worsening Across the Country*, PEN AMERICA (Dec. 13, 2021), https://perma.cc/8M74-64MB.

[3] *See* KNIGHT FOUNDATION & IPSOS, FREE EXPRESSION IN AMERICA POST-2020: A LANDMARK SURVEY OF AMERICANS' VIEWS ON SPEECH RIGHTS 21–22 (2021), https://perma.cc/R4SK-ANQL.

[4] *See* CATO INSTITUTE, CATO INSTITUTE FREE SPEECH AND TOLERANCE SURVEY (2017), https://

- 53% of Republicans favored stripping U.S. citizenship from people who burn the American flag;
- 51% of Democrats supported a law that would require that people refer to transgender persons by their preferred pronouns;
- 47% percent of Republicans favored a ban on the construction of new mosques;
- 58% of Democrats reported that employers should discipline employees for offensive Facebook posts.

Toleration has always been a virtue more readily extended to the in-group than the out-group (with principled actors few and far between). Still, the fact of the matter is that there is a widespread perception that the speech environment on the whole has become less free in recent years. To take just one example, a Knight Foundation survey found that just 47% of college students feel that their freedom of speech is secure, compared to 73% in 2016.[5] Notably, our legal environment for free speech is little changed. What does appear to be different is the degree to which speech is met with peer punishment and calls for corporate suppression.

In view of these developments, it is worth asking whether calls for the suppression of speech reflect wisdom or folly. Answering this question in turn requires asking (1) what motivates persons to calls for the private or public suppression of speech? It also requires asking (2) what do we know about the suppression of speech generally, independent of the pressures of the particular moment?

Regarding (1), I claim, in line with much work on free speech, that calls for censorship typically target speech judged to be harmful or dangerous by the persons issuing the calls. People aim to suppress speech because they worry about the effects of tolerating it. The precise nature of the concerns will vary by the case. Regarding (2), liberal theory provides reasons for worrying about calls of this sort.

One worry is that calls for censorship and intolerance of these kinds will overreach, targeting speech that is perfectly above-board. For all we know the propositions suppressed might turn out to be true or partially true in ways that advance the

---

perma.cc/97QN-KDCP.

[5] KNIGHT FOUNDATION & IPSOS, COLLEGE STUDENT VIEWS ON FREE EXPRESSION AND CAMPUS SPEECH 2022: A LOOK AT KEY TRENDS IN STUDENT SPEECH VIEWS SINCE 2016, at 3 (2022), https://perma.cc/KTH4-GHZ6.

conversation.

A second worry is that, even if the speech directly targeted is genuinely false and offensive, calling for its suppression can lead others with acceptable views to withdraw from the conversation for fear that they too will find themselves in the crosshairs. And in large societies there is good reason to worry about incentivizing people to hide their considered judgments from public view.

Third, even if all of the speech both targeted and affected by calls for suppression is genuinely bad considered in isolation, the net benefits of tolerating it might outweigh its costs in the long run. After all, engaging with false speech properly trains our rational faculties and puts us in a better position with respect to the grounds for belief.[6] It can also spur creativity and put us in the position of new, better arguments, and help us develop new, better ideas.

These three points encompass the main of the canonical Millian case in favor of free speech and against suppression by state, society, and social groups.[7] As a sociological matter, many of those who accept these arguments as grounding sound principles for toleration by state and non-state agents happen to be skeptical that the speech targeted is especially dangerous or that it rises to the level of harm.[8] To be sure, some see value in tolerating speech that rises to the level of significant harm. But many believe that harm demarcates the sphere of permissible intervention. Thus many disagreements about the scope of the free speech principle come down to disagreements about what speech is harmful and how harmful it is. Those convinced that a certain class of speech results in significant harm will be, by the lights of this dominant theory, rightly skeptical that these typical liberal arguments are applicable to them.

My goal in this paper is to argue that those most concerned about the *negative* effects of a certain class of expression should be especially reluctant to engage in public calls for its suppression. I will say that one publicly calls for the suppression of speech when one publicly demands its removal or sanctioning (or deploys

---

[6] *See* JOHN STUART MILL, ON LIBERTY 103 (D. Bromwich et al. ed., Yale Univ. Press 2003).

[7] *See id.*

[8] For example, in his entry on free speech in Stanford Encyclopedia, van Mill notes that the Nazi march through Skokie, Illinois resulted in much offense and outrage. But he questions whether those offended and outraged were harmed. See David van Mill, *Freedom of Speech*, § 2.3, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2021), https://perma.cc/B5H2-L8FA.

sufficiently strong sanctions directly targeting speech on the grounds that it deserves suppression). Relevant sanctions might include demanding that social media platforms remove or deboost it (by the social media platform or by another party), or that a person who utters some class of speech be fired. Public calls for suppression may also take the form of agitating for banning the relevant speech through legislative means or expansion of tort liability, or threatening to sue or demanding its removal under existing law.

In arguing that such public efforts to censor are misguided insofar as they target harmful speech, my argumentative strategy is aimed only at those who agree that the relevant speech is harmful. But it is hard to imagine that those that *do not* so believe will have much reason to engage in the kinds of public calls I have in mind. After all, they are not (typically) worried about the effects of the speech in question.

I begin in Part I with an account of bad speech and explain why its propensity to harm leads many people to want to suppress it. I then argue in Part II that there are three mechanisms that make public calls for such suppression themselves dangerous, such that those inclined to call for the suppression of speech typically undermine their own aims. The result, I think, is that we should think of calls to suppress even bad speech as themselves further instances of bad speech (i.e., speech against which responsible speakers ought to recognize a presumption). In Part III, I respond to objections to this seemingly paradoxical conclusion.

I. BAD SPEECH

Despite partisan differences regarding which speech is considered worthy of suppression, those wishing to suppress speech do so for a reason. They might seek to suppress because they believe that the speech's spread will harm their material interests. They might do so because they believe that exposure to some speech is threatening to some political, moral, or religious orthodoxy. Or they might do so because they believe that the speech is harmful: that it wrongfully sets back the interests of some person or group of persons or otherwise undermines some value (say, public health or safety) that we have reason to care about.

Here are some examples to illustrate this. The broadcaster that suppresses a sitcom's satirical representation of Chinese censorship because such criticism will alienate Chinese viewers hopes to profit thereby.[9] The parent that calls for her child's

[9] *See* Emily Nussbaum, *CBS Censors "The Good Fight" for a Musical Short About China*, NEW YORKER (May 7, 2019), https://perma.cc/633B-5598.

Catholic school to ban books that glorify lives of sin hopes to ensure that her child's education does not lead her off the true path. The group that calls for a press to rescind the publication of a book whose author argues in favor of legislatively protected spaces reserved for biological females worries that this might set back the interests of trans persons (either by depriving them of space to which they have a claim to enter or by fueling anti-trans sentiment and anti-trans violence). The person anxious about COVID-19 who calls for social media platforms to remove content questioning the efficacy of vaccines hopes to stop speech which she believes compromises public health. The conservative who calls for publications to take a stand against critical race theory seeks to protect her children from the self-loathing that she fears will follow from too heightened an awareness of racial politics at a young age or perhaps from indoctrination into an ideology that she finds controversial.

I suspect that examples like these are familiar to most readers. In the background is a view of speech according to which its effects extend beyond the expression of propositions. Speech itself can and regularly does issue in harm, understood as a setback to a person's, group's, or community's interests.[10] When the harms reach a certain threshold, the behavior that issues in them admits of social regulation and (in extreme cases) political or legal regulation. It would be nice if speakers self-regulated and exercised their free speech rights responsibly. Unfortunately, this is seldom the case and so we need to find effective means of policing speech. If state remedies are off the table, then we'd better look for social mechanisms to punish speech. So-called "cancel culture" (or, if you prefer, accountability culture)—whether in its right- or left-wing flavors—marks just one example of an attempt to

[10] Well-studied classes of harmful speech include (but are not limited to) hate speech, *see, e.g.*, JEREMY WALDRON, THE HARM IN HATE SPEECH (2014); NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP (2018); sexist speech, *see, e.g.*, Rae Langton, *Speech Acts and Unspeakable Acts*, 22 PHIL. & PUB. AFF. 293 (1993); Mary Kate McGowan, *Oppressive Speech*, 87 AUSTRALASIAN J. OF PHIL. 389 (2009); MARY KATE MCGOWAN, JUST WORDS: ON SPEECH AND HIDDEN HARM (2019); Ishani Maitra, *Silencing Speech*, 39 CANADIAN J. PHIL. 309 (2009); misinformation, *see, e.g.*, CASS R. SUNSTEIN, LIARS: FALSEHOODS AND FREE SPEECH IN AN AGE OF DECEPTION (2021); Hunt Allcott, Matthew Gentzkow & Chuan Yu, *Trends in the Diffusion of Misinformation on Social Media*, 6 RESEARCH & POL. 1 (2019), https://perma.cc/ZQ8J-837H; and ideological speech or propaganda, *see, e.g.*, NOAM CHOMSKY, NECESSARY ILLUSIONS: THOUGHT CONTROL IN DEMOCRATIC SOCIETIES (1989); YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (2018).

do just this.

Of course, many deny that speech can have these negative impacts. They will point to the fact that the effects of speech vary considerably and are mediated by a culture's approach to the harmfulness of speech. (The more speech is taken to harm, the more it in fact harms.[11]) They might point to the benefits of developing a kind of resilience to affronts that is possible only through adversity.

Others admit that speech can harm but are well-convinced that restricting harmful speech is on balance more harmful than tolerating it. They might point out that protecting people from misinformation relies on the assumption that people are not capable of assessing information for themselves. Even if this assumption is true, they will say, it is by no means clear that powerful elites can be trusted to protect persons from themselves without risking abuse. (Might the elites use the mandate to hide dangerous information from view to instead conceal information that might compromise their power?) And even if such elites succeed in targeting only bad speech, still many will worry that our outsourcing the tasks of democratic self-government will in the long term undermine our capacities to think critically.

I do not wish to gainsay these arguments. They have a venerable history. And if they depend for their success on empirical questions that are beyond the competence of a philosopher to answer, those questions remain controversial among experts. My point here is more modest, namely to limit the scope of my argument. Persons who accept this family of arguments or for other reasons think that restricting speech will on balance cause more harm than good will see little reason to suppress speech and so there is little point in convincing them that they ought not to.

For that reason, I shall assume for the purposes of this paper what I really believe, namely that it is sensible to be concerned about the harms of speech—no less sensible, in any case, than it is to be concerned about the harms of silence. Moreover, sometimes the effects of harmful speech rise to a level at which it is sensible to intervene to suppress it. Taking this much for granted, I argue that those who see things this way should *also* see calls to suppress harmful as themselves a class of bad speech, *i.e.*, speech from which responsible speakers should refrain. If there is a duty to refrain from harmful speech, then there is a duty to refrain from publicly

---

[11] *See, e.g.*, April Bleske-Rechek et al., *In the Eye of the Beholder: Situational and Dispositional Predictors of Perceiving Harm in Others' Words*, 200 PERSONALITY & INDIVIDUAL DIFFERENCES 111902 (2023).

calling for the suppression of speech.

So far, we have seen that there are reasons to worry that speech can exert a pernicious impact on our lives (epistemic and otherwise). This explains, at least to a substantial degree, impulses to suppress or punish certain speech acts. Those who wish to suppress speech do so because of a shared view that speech can harm, though there is a diversity of views concerning which speech rises to a level of harm sufficient to warrant suppressing it.

Though they disagree on these particulars, it is important to note that I think they agree that the mechanism by which speech harms is *exposure*.

Exposure Principle: Bad speech harms in virtue of its effects on those exposed to it.

Why should we believe this principle, aside from its facial plausibility? Well, for one, it's hard to see how speech could harm if no one were exposed to it. A genocidal manifesto might be repugnant, but if it is buried under yards of unturned earth in the middle of nowhere, it fails to harm. If we still think of it as harmful, this is because we imagine its effects if discovered or on the writer or past readers, not because of its properties when lonely under dirt.

By contrast, exposure makes uptake of misinformation or hateful attitudes or sin possible; it makes it possible for those exposed to sense a set-back to their dignitary or other interests. This is exactly why people seek to suppress (*i.e.*, limit the exposure to) bad speech. A corollary of the exposure principle is that, as more people are exposed to the relevant speech, more people are potentially harmed by it, either because they act in line with it or revise their beliefs (potentially about themselves) in misguided ways in response to it.

II. Calls for Suppression as Bad Speech

If we ought to refrain from speech, exposure to which causes harm, then there is reason to believe that we ought also to refrain from calls to suppress such speech. Or so I will argue in this section.

A first argument for this paradoxical conclusion can be summed up in the following five steps.

(1) Bad speech is speech, exposure to which causes or constitutes harm.

(2) If speech augments exposure to bad speech, then that speech causes or constitutes harm.

(3) Calls to suppress bad speech augment exposure to bad speech.

(4) Therefore, calls to suppress bad speech cause or constitute harm.

(5) Therefore, calls to suppress bad speech are bad speech.

Premise (1) sums up the lesson from the last section. Premise 2 is supposed to be read in light of the exposure principle. Recall, according to the exposure principle, bad speech works its pernicious effects through exposure. Thus, for each person who interacts with the speech, there is some nonzero probability of harm to the new listener or the new listener's causing or constituting harm to some non-listener.

Here are some examples. As racist hate speech spreads, more are exposed to it and it causes greater harm to that degree, either by inciting violence against its targets or setting back their dignitary interests or compromising their self-respect. As misinformation proliferates, false beliefs, imprudence, and dangerous behavior follow. As ideological speech becomes more prevalent, more people will drink the Kool-Aid. As sexist speech becomes common, women's complaints fall more frequently on deaf ears or their speech is not taken seriously or they are more often victims of violence. And so on. Generally: As some harmful speech reaches a larger audience, we should expect further harm to result, more or less mechanistically.

So much for premise (2). (4) follows necessarily from (2) and (3). (5), the final conclusion, follows straightforwardly from (1) and (4). If this accounting is right, the success of the argument turns on premise (3), which is controversial. In the remainder of this section, I will describe three mechanisms through which calls to suppress speech paradoxically increase exposure.

First, and perhaps most familiarly, there is the Streisand effect—named after Barbara Streisand's infamous attempt to suppress images of her home placed on the internet by the California Coastal Project. Before Streisand sued the photographer, the image had received a total of six downloads. A month later, after Streisand's lawsuit made the news, the image was accessed over four hundred thousand times.[12] While the Streisand effect is not inevitable, it is likely to obtain in media environments where outlets profit from publicizing a person's embarrassment over the existence of some information or when suppression of speech is newsworthy for other reasons. Naturally, if the would-be censors better understood their information environment and what their censorship signaled, then they would not call for the suppression of the content in the first place. Unfortunately, people are only

[12] *See* Christian Gläßel & Katrin Paula, *Sometimes Less Is More: Censorship, News Falsification, and Disapproval in 1989 East Germany*, 64 AM. J. POLI. SCI. 682, 682 (2019).

partially sophisticated. They tend to overlook that people treat a desire to suppress information as evidence that it is interesting or embarrassing.[13]

Consider a recent open letter on behalf of the USA branch of the Oxford University Press Guild demanding that Oxford not publish Holly Lawford-Smith's forthcoming title *Gender-Critical Feminism*.[14] The open letter failed to convince editors at the press to withdraw their commitment to publish. What's more, it garnered substantial attention on social media, bringing the book and its general orientation more readily into the public view. Now, the call to suppress the publication implies that it is a problem for the book to reach an audience of, say two hundred people (which, according to a recent poll,[15] is the average readership for academic books). But, again, the call failed and drew significant attention to the book. One forum responded to the calls for suppression by advertising a 30% off sale and encouraged readers to buy the book. The Streisand effect implies that we should expect the various open letters enjoining OUP to censor the book to increase its audience and therefore its exposure. The more harmful exposure is, the greater the risk of open letters like this.[16]

The Streisand effect is the most direct way in which calls to suppress speech can increase exposure. It functions by literally drawing attention to the relevant

---

[13] For discussion and attempts to measure, explain, and model the effect, *see generally* Sue Curry Jansen & Brian Martin, *Making Censorship Backfire*, 7 COUNTERPOISE 5 (2003); Zubair Nabi, ~~*Resistance*~~ *Censorship is Futile*, 19 FIRST MONDAY (2014); Sue Curry Jansen & Brian Martin, *The Streisand Effect and Censorship Backfire*, 9 INT'L J. COMM. 6561 (2015); Jeanne Hagenbach & Frédéric Koessler, *The Streisand Effect: Signaling and Partial Sophistication*, 143 J. ECON. BEHAV. & ORG. 1 (2017); Sujay Bhatt & Tamer Basar, *Streisand Games on Complex Social Networks*, 2020 59TH IEEE CONF. ON DECISION & CONTROL (CDC) 1122 (2020).

[14] *See* @oupusaguild, TWITTER (April 11, 2022, 7:10 PM), https://perma.cc/5D58-DYWZ (the petition by OUP USA Guild was taken down after the press formally responded). This open letter is not to be confused with the letter from OUP affiliated academics who merely ask for an accounting of what steps OUP took to ensure the rigor of Lawford-Smith's book and the steps the press planned to take to mitigate any harm that it might cause. For that letter, see *Letter to OUP re: "Gender Critical" Publication*, https://perma.cc/F2WQ-MDTD.

[15] Donald A. Barclay, *Academic Print Books Are Dying. What's the Future?*, THE CONVERSATION (Nov. 15, 2015, 5:45 AM), https://perma.cc/X8MK-UPCE.

[16] Another example: Anderson Cooper negatively covered the subreddit "jailbait." After he did, traffic to that particularly depraved corner of the internet quadrupled. *See* ANDREW MARANTZ, ANTISOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS, AND THE HIJACKING OF THE AMERICAN CONVERSATION 211 (2020).

content, sometimes on a broad scale, and broadcasting its existence to those who would have otherwise remained ignorant of it. Moreover, by signaling that the information is embarrassing or inconvenient or disfavored, the call for suppression generates in audiences a natural desire to put in some effort toward uncovering it. Media, driven by incentives to profit, draw attention to the act of censorship, and people respond by seeking it out.

A second mechanism concerns uptake. Calls to suppress speech can increase the risks of exposure without increasing audience size by making people take the ideas seriously when they otherwise would not do so.[17] Psychologists call the general phenomenon *reactance*. Reactance consists in responding to the denial of one's freedom (in this case, to say, think or believe certain things) through a reassertion of freedom. If reactance is a common response to calls for censorship, we should worry that the latter might backfire by making listeners more likely to believe, engage with, or platform the problematic speech. Not only do people seek out *exposure* to censored speech (the Streisand effect), but some will treat the act of censorship as offering a *reason* for belief formation and expression against the efforts of the censor.[18] Whereas the Streisand effect makes exposure to putatively bad speech more likely, reactance suggests that bad speech will enjoy a higher probability of uptake among those exposed to both the speech and the call to suppress it than to the speech if left alone.

The reactance effect is well studied by political psychologists. Nearly sixty years of research in social psychology and communications finds that prohibiting something or attempting to control others' behavior leads to increased motivation to engage in the behavior.[19] And scholars have not missed its implications for the

---

[17] *See generally* Stephen Worchel & Susan E. Arnold, *The Effects of Censorship and Attractiveness of the Censor on Attitude Change*, 9 J. EXPERIMENTAL SOC. PSYCH. 365 (1973); Stephen Worchel et al., *The Effects of Censorship on Attitude Change: The Influence of Censor and Communication Characteristics*, 5 J. APPLIED SOC. PSYCH. 2279 (1975).

[18] A distinct, but related, problem is that when audiences suspect censorship norms are constraining the things that people say in public, they are less likely to believe that speakers are being sincere. *See generally* Lucian Gideon Conway et al., *When Self-Censorship Norms Backfire: The Manufacturing of Positive Communication and Its Ironic Consequences for the Perceptions of Groups*, 31 BASIC & APPLIED SOC. PSYCH. 335 (2009).

[19] The seminal study is JACK BREHM, A THEORY OF PSYCHOLOGICAL REACTANCE (1966). *See* also Benjamin D. Rosenberg & Jason T. Siegel, *A 50-Year Review of Psychological Reactance Theory: Do Not Read This Article*, 4 MOTIVATION SCI. 281 (2018). For recent applications in domains as

efficacy of censorship and suppression.[20] While the precise ground of reactance is unclear, it isn't hard to imagine that rational agents view calls for censorship or suppression as a tacit admission that the person making the call is not in possession of a good argument against the maligned position.

Public calls for censorship (along with actual threats in response to speech) are highly likely to elicit a reactance response. To see this, consider three reactance conditions that are relatively common in the political psychology literature:

> First, a decision maker must believe that they hold decisional autonomy within a specific domain . . . . Second, psychological reactance requires a stimulus. It arises when another actor directly acts to impinge on the decision maker's autonomy . . . . Third, forceful, dogmatic language intensifies reactance . . . . Message features amplify or mitigate reactance by altering the degree to which a target perceives it as threatening their autonomy.[21]

Note now that each condition is met in the cases at hand. Because people widely believe that they ought to be free to speak and listen without outside constraint, they believe that they hold decisional autonomy within the domain of what is to be heard. Because public calls for the suppression of speech would undermine listeners' access to speech by limiting speakers' ability to express the speech, they clearly involve an actor directly acting to impinge on both listeners' and speakers' autonomy. Finally, because they aim to prevent harm, such calls are (though contingently) often made forcefully and dogmatically.

Here are some examples that illustrate the mechanism. First, in his investigative report on the Alt-Right, the journalist Andrew Marantz tells the story about Cassandra Fairbanks' radicalization from a disaffected moderate to a full-on Trump supporter. She began by testing the waters, publishing Tweets that she describes as "not completely anti-Trump." The response was, as she describes it, hysterical. "I got called a literal Nazi so many times, I eventually went Fuck it, I'll just go all in."[22] At that point, she stopped writing for liberal outlets and began writing explicitly

---

significant as international politics, *see* Kathleen E. Powers & Dan Altman, *The Psychology of Coercion Failure: How Reactance Explains Resistance to Threats*, AM. J. POL. SCI. (forthcoming 2022), https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12711.

[20] For a recent example, *see generally* Golnoosh Behrouzian et al., *Resisting Censorship: How Citizens Navigate Closed Media Environments.*, 10 INT'L J. COMM. 4345 (2016).

[21] *See* Powers & Altman, *supra* note 19, at 6.

[22] *See* MARANTZ, *supra* note 16, at 12.

pro-Trump pieces. Her thinking seemed to be: Look, if there is no rational response to what I'm doing and instead people are trying to shame me out of these views (which I am free to form), I may as well take a stronger stand and see what happens.

Or consider the way Mike Cernovich, a celebrity of the Alt-Right, describes a speech by Hillary Clinton, in which she targets a number of Breitbart headlines for being beyond the pale. Clinton's speech, he says, "was the stupidest thing she could have done" insofar as it reveals the degree to which Clinton's social media team was "triggered" by the headlines and "asked their boss to yell at us and make us go away."[23] He continues: "Well, we're not going away. They just made us stronger."[24] Mike Enoch tells a similar story: When he began trafficking in race science, he was initially put off by the attempts to silence him by calling him a racist. But then he realized that that this was a mere emotional appeal. He never met reasoned responses to his arguments, only a series of claims that what he was saying was beyond the pale; this motivated him to go further, rather than to stop.[25]

A common explanation of behavior like this is that human beings have an interest in freedom that leads them to bristle when they believe that others are trying to compel their assent where it might not naturally lead. If so, we should expect calls for the suppression of speech to be especially dangerous in cultures that prize autonomy.[26] Another (compatible) possibility is that there is simply something attractive about forbidden fruit.[27] Describing an incident in which he and some friends stole pears, Augustine famously confesses his motivation was "merely the excitement of thieving and the doing of what was wrong;" criminality was, by Augustine's

---

[23] *Id.* at 174–75.

[24] *Id.* at 180.

[25] *Id.* at 297–98.

[26] This is the general reasoning offered by psychologists studying reactance. *See* Christina Steindl et al., *Understanding Psychological Reactance: New Developments and Findings*, 223 ZEITSCHRIFT FÜR PSYCHOLOGIE 205 (2015), for two recent reviews of the literature.

[27] Brad J. Bushman & Angela D. Stack, *Forbidden Fruit Versus Tainted Fruit: Effects of Warning Labels on Attraction to Television Violence*, 2 J. EXPERIMENTAL PSYCH.: APPLIED 207 (1996), for instance, find that warning labels from authorities against violent television content are met with increased interest in such content. By contrast, merely informative messages about the content do not have this effect. Additionally, a sign in a restroom prohibiting graffiti led to increased graffiti, especially when the prohibition came from a recognized authority figure.

lights, the "piquant sauce" that made the act of taking *these* pears delightful.[28]

Additionally, if one perceives that one's political opponents believe one's speech dangerous, that is pretty near a reason to engage in it, if for no other reason than to signal one's fealty to the cause. A final possibility is that calls for suppression increase the appeal of propositions because censorship tends to be most attractive when one does not possess a compelling proof. Whatever the explanation, we should worry about calling to suppress bad speech insofar as doing so increasingly disposes speakers and listeners favorably to the relevant (and by hypothesis, harmful) speech.

Both the Streisand effect and the psychological phenomenon of reactance fit more or less neatly into the argument above.[29] But there is a more direct way that calls to suppress harmful speech can themselves harm. Consider the following argument:

(1) Bad speech is speech, exposure to which causes or constitutes harm.

(2) Public calls for censorship (a kind of speech) enable unscrupulous speakers to claim a kind of victim status.

(3) Speakers who can claim victim status enjoy augmented social power.

(4) Therefore, public calls for censorship enable unscrupulous speakers to have more social power.

(5) Enabling unscrupulous speakers to have more social power causes or constitutes harm.

(6) Therefore, public calls for censorship cause or constitute harm.

(7) Therefore, public calls for censorship are bad speech.

Once again, premise (1) is the result of the previous section. Premise (2) is controversial and we shall have occasion to defend it. Premise (3) is plausible (we shall

---

[28] AUGUSTINE, CONFESSIONS 29–31 (H. Chadwick ed., Oxford University Press 2008).

[29] One might object that increased uptake need not cause increased exposure and so requires a separate treatment. But as more people develop sympathy with the ideas expressed in the speech that is by hypothesis beyond the pale, there are that many more people who might express the ideas or otherwise act upon them. By contrast, there is a sense in which if everyone exposed to bad speech were indifferent to it, many of its worst effects would not obtain. Those exposed to misinformation would not be drawn in; those exposed to hate speech would see it as small-minded foolishness and would not take it to legitimate oppression of its targets; those exposed to ideological propaganda would recognize it as such and move on accordingly.

see) in view of recent work in sociology and social psychology. Step (4) follows deductively from (2) and (3). Premise (5) is plausible, though not certain. It is of course possible that unscrupulous persons would never use their social power to harm. But given that they are unscrupulous (*i.e.*, lacking in scruples), such a position is not plausible. The sub-conclusion (6) follows from (4) and (5). Finally, the main conclusion (7) follows validly from (1) and (6).

If this accounting is right, the main premises to investigate are premises (2) and (3). Beginning with (2), why might public calls for censorship enable targets to claim a kind of victim status? The basic idea is that, especially where rights to speak are broad and people prize an environment in which expression is free, the claim that one's perspective is being suppressed garners considerable sympathy from disinterested third parties. Even if the call to suppress speech is directed at private actors who cannot violate free speech rights held against the state and guaranteed by the constitution, many individuals are confused about this and think that private suppression of speech violates rights. Even if they are not confused on this constitutional point, it is widely recognized (recent survey data notwithstanding) that there is considerable value in tolerating a range of speech that appears distasteful.

Accordingly, when persons claim that their speech has not enjoyed such toleration, this can motivate others to rally behind them. "I don't agree with the content of this speech," we might imagine them saying, "but it deserves to be heard." Note for instance that researchers find that persons who wish to engage in hate speech frequently invoke their rights to speak and express themselves when social pressure is brought to bear on them. They likely do this in part because it is effective in defusing threats to their ability to express themselves and in gaining sympathy from third parties.[30]

But, so what if speakers are able to gain such sympathy? Why think that this will lead to undesired effects? The idea, expressed in premise (3) above, is that this sympathy for victims whose widely recognized social rights are perceived to be violated more or less directly leads to increased social power.[31] Even if a speaker's

---

[30] Mark H. White & Christian S. Crandall, *Freedom of Racist Speech: Ego and Expressive Threats*, 113 J. PERSONALITY & SOC. PSYCH. 413, 424 (2017).

[31] *See, e.g.*, BRADLEY CAMPBELL & JASON MANNING, THE RISE OF VICTIMHOOD CULTURE: MICROAGGRESSIONS, SAFE SPACES, AND THE NEW CULTURE WARS 106 (2018); see also the growing literature on competitive victimhood (e.g., Rotem Kahalon et al., *Power matters: The role of power and morality needs in competitive victimhood among advantaged and disadvantaged groups*, 58 British

ideas are hated, the steadfast belief that free speech requires that they be heard (and that this value is being compromised by the censors) can motivate others to offer them new platforms in virtue of their having lost old ones (*e.g.*, new book deals and speaking engagements). If the call to suppress speech occurs in a polarized environment and is supported by only one political team, this can vault the speaker into a position of high status on the opposite team. To the degree that the speech suppressed was genuinely harmful, this augmentation of social power better positions the speaker to act on his or her ideas. But if those views are bad, then putting them into action is presumably worse still.

I do not mean to suggest that these are the only mechanisms that tell against publicly calling for censorship. There may well be others. But I do take it that these present powerful reasons to think that public calls to suppress speech will themselves be harmful.

It would be a mistake to infer simply that engaging in public calls to suppress speech might be sometimes counterproductive. Rather, we should conclude something stronger. Insofar as the targeted speech is harmful in the ways suggested above, and insofar as calls to censorship increase exposure to that speech, increase the probability of its being taken up by listeners, and increase the social power of those who utter the speech, public calls to suppress bad speech are *themselves* bad speech. After all, if bad speech is bad in virtue of its consequences, and if public calls to censorship can have similar consequences, calls to censorship too should count as bad speech.

III. Objections and Replies

So much for my basic argument. I anticipate a number of objections, which I do my best to answer in this section.

According to a first objection, bad speech is not bad and *prima facie* worthy of suppression merely because of its consequences, but owing also to its intent (in cases where the agent seeks to harm) or failure of due care (in cases of negligent and reckless speech). Bad speech worthy of the name operates in bad faith. It is not merely that it misleads and marginalizes, but that it is designed to do so or that it does so in disregard or culpable ignorance of its effects. But because public calls to censor are themselves aimed at preventing rather than causing these effects they do not count as bad speech in the same way.

Journal of Social Psychology 452 (2019)).

This objection will not do. After all, many of those who hold views that others wish to suppress are acting in good faith and are just badly confused about the true, the right, and the good. Many understand that some speech is irresponsible, have done their due diligence, and have come to the wrong conclusion. There is no question, for example, that there are good faith believers in all manner of conspiracy theories.

What's more, the objection tacitly admits that good-faith bad speech ought to be tolerated. In turn, this means that we would have to know whether a person is acting in good faith to justify calling for speech suppression. But our intentions are difficult to discern. We do not wear them on our sleeves. If justifying calls to suppress speech depends on an estimation of a person's intentions, these estimations are fraught.

Finally, calls for censorship themselves need not be aimed at salutary ends. Many people call for censorship to gain approval from their in-groups or to otherwise increase their social standing.[32] If so, even if we grant the objection, those calls to suppress speech that are motivated by status-seeking and other non-benevolent aims will still count as bad speech, as I allege above.

The second objection observes that none of these mechanisms is guaranteed to operate with respect to a given call to suppress bad speech. Some such calls might succeed in avoiding these kinds of effects and succeed in suppressing the targeted speech.

I do not, naturally, wish to deny this obvious fact. Some calls to suppress speech succeed. If they didn't, philosophers like Mill and Tocqueville would've been foolish to worry about the tyranny of the prevailing opinion and the intolerance of majorities in regimes that protect the freedoms of speech and association. Still, I want to say three things in reply.

First, no class of bad speech is *necessarily* harmful. Some misinformation is readily corrected in a way that leaves all parties to the discussion better off. In the right social circumstances, some hateful speech causes protest and counterspeech that leads to changes that better uphold the rights of minorities than the previous order. And some sacrilege strengthens faith. What opponents of each kind of bad speech claim should not, then, be that these classes of speech are necessarily

---

[32] *See, e.g.*, JUSTIN TOSI & BRANDON WARMKE, GRANDSTANDING: THE USE AND ABUSE OF MORAL TALK (2020).

harmful on net, but rather that they tend toward being so. Thus, as long as public calls to censorship *tend* to have the effects that they have, they are on all fours with the kinds of bad speech they target.

Second, even if there *were* some classes of necessarily harmful speech, those calling for the suppression of that speech should be concerned about taking any action that has a substantial probability of exacerbating the relevant harms. And even if I have stopped short of showing that public calls to censorship certainly exacerbate harms, I think I have done enough to show that they raise the probability sufficiently to create a presumption against them, one that might be overridden in the right circumstances, but that imposes a duty of due diligence on those who issue them. I think it's clear (though I cannot argue for this here) that such calls are frequently issued in violation of such a duty.

For that reason, third, it is important not to read me as making a stronger claim than I wish to make. I am certainly not saying that public calls to suppress speech are necessarily harmful or unjustified in all cases. I am only arguing that there are general reasons to think that such speech will be harmful in the same way other bad speech is harmful. But how to exercise one's speech rights responsibly is a difficult question, and my task is not to deliver an answer adequate to all eventualities. In any given case, there are going to be tradeoffs and probabilities to assess before determining what finally to do. These tradeoffs should be made in full awareness of the likely consequences of our alternatives. When a clear-eyed assessment along these lines favors calling to suppress speech, that particular call is unlikely to be an instance of bad speech.

A final objection is more conciliatory. It grants that calls to suppress speech might be counterproductive in their direct effects but denies that these direct effects are the only thing that matters. Calls to suppress speech might be motivated by their anticipated indirect effects, for example their ability to signal to bystanders the bounds of acceptable speech, to express solidarity with victims of bad speech, to try to found a new norm according to which the relevant speech is more widely accepted as beyond the pale, and to communicate higher-order evidence about what to believe.[33]

I do not deny that public calls to suppress speech can have these goals. But each goal admits of pursuit by means other than *publicly demanding the suppression of*

[33] Neil Levy, *Virtue Signalling Is Virtuous*, 198 SYNTHESE 9545 (2020).

the speech. Applying social pressure, expressing indignation, explaining what's wrong with the speech, arguing in favor of new norms: These are all ways of achieving these goals that are less subject to the mechanisms discussed above. For that reason, they may present better strategies. At the end of the day, however, what strategies work best to realize which ends is an empirical question in need of further investigation.

Conclusion

Liberal norms and institutions have faced distinctive challenges in recent years. These challenges have made clear the degree to which free speech rights can be exercised poorly to the detriment of us all. Even if we do not wish to respond to these problems by curtailing constitutional protections for free speech and assembly, it can nevertheless seem as if we must do something. And engaging in the public attempts to suppress speech by nongovernmental means can seem an attractive way of keeping speech from issuing in harm without violating or dangerously limiting anyone's constitutional rights. They are, indeed, *expressions* of those very rights.

I have argued in this paper, however, that calls to suppress bad speech ought to be understood as themselves a class of bad speech. This is because there are widely operating mechanisms according to which calls to suppress speech can in fact increase its reach or power. They can increase the reach of bad speech insofar as they increase exposure to it by virtue of bringing the bad speech to the attention of those who might not otherwise have seen it. They can increase the negative power of such speech by increasing the probability that it enjoys uptake. And they can more or less directly increase the social power of their targets.

I hasten to add that this is not so much an argument against censorship as it is an argument against publicly calling for it. It is also, I think, an argument against *failed* attempts to censor, which may be more or less covert in design but come to public light.[34] But for all that, if one can genuinely stop the distribution of harmful

---

[34] Yevgeniy Golovchenko, *Fighting Propaganda with Censorship: A Study of the Ukrainian Ban on Russian Social Media*, 84 J. POL. 639 (2022), finds that transparent censorship can, under the right conditions, effectively reduce exposure—in particular, that Ukrainian censorship of a Russian social media platform subject to Russian surveillance substantially reduced traffic to it, in spite of there having been ways around the censorship (*e.g.*, easy VPN access). But it is not clear by Golovchenko's own admission that these findings are generalizable. For instance, it seems plausible that part of the ban's success was access to substitutes (in this case, Facebook) to the banned website (*id.* at 651). When censorship bans content rather than a platform, substitutions are unlikely to be available.

speech, then persons will not happen upon it. In the limit, successfully suppressed speech will be no more dangerous than a racist manifesto buried deep in the ground. In more realistic cases, the effects of suppression are simply of a greater magnitude than any of its amplifying effects. If what I have argued is correct, it will be difficult to predict when this will be so. But when it is so and when we have good reason to believe that it is, the arguments in this paper say nothing against suppression. But to qualify my argument in this way is not to recommend covert censorship. After all, the risks of exposure are high, as are the legitimacy risks of institutions that operate by subtly suppressing dissent and deviant speech.

To the degree that Americans are right to worry about bad speech and to the degree that our culture wars are, to a substantial degree about how to handle it without state intervention, then, I believe that we must look beyond calls for private suppression and attempts to dial up social sanctions for such speech. We must be creative in seeking alternatives and bear in mind that the actual effects of our interventions do not always match the effects we intend to bring about. In our endeavors to better calibrate our responses in view of the many things that we care and ought to care about, we should think hard about their probable effects given a diverse group of observers.

Calls to censor bad speech are content-based in precisely this way. *See also* William R. Hobbs & Margaret E. Roberts, *How Sudden Censorship Can Increase Access to Information*, 112 AM. POL. SCI. REV. 621, 621 (2018).