FREE EXPRESSION IN THE AGE OF SOCIAL MEDIA

INTIMIDATION CULTURE: FREE EXPRESSION IN THE AGE OF SOCIAL MEDIA

By FRANCESCO STELLIN STURINO, M.A., B.A.

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2025) Hamilton, Ontario (Philosophy)

TITLE: Intimidation Culture: Free Expression in the Age of Social Media

AUTHOR: Francesco Stellin Sturino, M.A. (University of Western Ontario)

SUPERVISOR: Professor Violetta Igneski

NUMBER OF PAGES: ix; 251

Lay abstract: While social media platforms have dramatically bolstered the ability of ordinary people to broadcast their views to large audiences, the dynamics of online communication have also had a stifling effect on public discourse. Due to social media's tendency to reward content that is extreme and divisive, it is often the case that people with more moderate views engage in self-censorship and preference falsification in order to evade online backlash. This project deploys the philosophy of the seminal liberal thinker John Stuart Mill in order to examine the phenomenon of online intimidation culture and assess its pernicious impact on society. It is argued that despite the persistent toxicity of social media discourse, the societal harms that it produces can be mitigated through the cultivation of institutions that are resilient in the face of pressure campaigns, and firmly committed to intellectual diversity and freedom of expression.

Abstract: While social media platforms have dramatically bolstered the ability of ordinary people to broadcast their views to large audiences, the dynamics of online communication have also had a stifling effect on public discourse. Due to social media's tendency to reward content that is extreme and divisive, it is often the case that people with more moderate views engage in self-censorship and preference falsification in order to evade online backlash. This project deploys the philosophy of the seminal liberal thinker John Stuart Mill in order to examine the phenomenon of online intimidation culture and assess its pernicious impact on society. Three social goods are identified that are jeopardized when thought and expression become constrained due to formal or informal censorship. These are critical intellectual faculties, authenticity in discourse, and equity in accountability. It is argued that those who are interested in preserving these social goods have strong grounds for resisting the pressures of intimidation culture and working to establish an atmosphere of free expression wherein people from diverse backgrounds can explore and assess a broad array of competing ideas without fear of punishment. It is likewise argued that despite the persistent toxicity of social media discourse, the societal harms that it produces can be mitigated through the cultivation of institutions that are resilient in the face of pressure campaigns, and firmly committed to intellectual diversity and freedom of expression. Finally, it is posited that social media is not inherently at odds with a Millian atmosphere of free expression. If the incentives that animate online discourse are realigned in order to encourage reasoned discourse rather than performative antagonism, then this technology could be an asset to humans' capacity for compassion by facilitating greater communication and understanding between individuals and groups from different parts of the world.

Acknowledgements: I am grateful to my PhD supervisor, Professor Violetta Igneski, for her guidance and support over the course of my doctoral studies. She has consistently encouraged me to pursue research questions in Philosophy that are innovative and relevant to contemporary issues, and this has been a valuable source of motivation for me. I also owe thanks to the two other members of my PhD supervisory committee, Professors Wil Waluchow and Stefan Sciaraffa, for taking the time to analyze my work and provide valuable recommendations about how it can be improved. In addition, I would like to extend thanks to McMaster University Philosophy Department Chair Mark Johnstone for his conscientiousness and professionalism, which I and many others aspire to emulate.

Table of Contents:

Declaration of Academic Achievement: The content of this dissertation is the original work of the author. Where the work of others has been quoted or referenced, citations have been included.

Chapter i: The Phenomenon of Online Intimidation Culture

i.i: The Currency and Incentives of Social Media

Social media platforms are undoubtedly among the most influential venues for expression that currently exist. Billions of people across the globe now use social media routinely, and the breadth of these platforms' reach continues to grow. As social media continues to influence nearly every facet of human life, from interpersonal relationships, to commerce, to the democratic process, it is vital that we reflect on the manner in which these platforms operate and the incentives that they introduce. If it turns out to be the case that our modern media ecosystem incentivizes people to behave poorly, then we may have grounds to seek revisions to it in the interest of realigning these incentives. This is especially, but not exclusively, true if and when media firms introduce incentives that motivate users to behave in a manner that is damaging to others. Insofar as we are interested in identifying and putting a stop to bad behaviour, we likewise must be attentive to the systems of informal reward and punishment that can animate such behavior.

Anyone who has experience using modern social media knows that these platforms provide users tools with which they can assess the extent to which their online content is successful in garnering attention from others. Every major social media platform that currently exists enables users to track the number of likes, shares, comments, etc. that a given piece of content has generated. These forms of interaction fall into the category of "engagement": they involve social media users choosing to interact with a piece of content that another user has posted. Some platforms even enable users to track the number of impressions that a given piece of content has generated. "Impressions" are simply views that a piece of content has received, without necessarily resulting in any specific form of interaction between users.[1] Engagement can plausibly be

[1] Normally, the content on social media receives far more impressions than engagement, as many social media users choose to refrain from interacting with content in the form of likes, shares, comments, and so on.

conceptualized as the "currency" of social media, as it signals to audiences that a piece of content is worth paying attention to, thereby driving further engagement. Sociologist Ilana Redstone and Brookings Institution senior fellow John Villasenor explain the nature of social media currency:

> Social media services are designed so that social media companies can get more traffic, users, and data, and, as a result, higher revenue and market value. As designers of social media services have long known, products that exploit (and contribute to) our distractibility by capitalizing on the human tendency to seek affirmation can be highly successful in the marketplace. For users, the currency of the realm in social media is likes, shares, comments, and retweets, which in combination satisfy a need for validation and attention.
> (2020, 34)

In cases wherein social media users are monetizing their content directly or indirectly, this social media currency can help generate real financial gains. Social media content that garners little or no engagement is generally overshadowed by other content that does a better job of securing the attention of users and motivating them to interact with the social media interface, which makes the social media environment competitive in an important sense.

In recent years, academics have been drawing attention to the role of social media in generating and exacerbating social tensions.[2] In order to understand this phenomenon, it is important to appreciate the significance of engagement in the realm of online discourse. Evidence indicates that if one has an interest in achieving popularity on a social media platform, one of the most reliable tactics for achieving this is to launch attacks on other individuals and groups in the online spaces that social media companies provide. In her book *How Civil Wars Start and How to*

[2] Examples include Stephen Macedo & Frances Lee (2025), Tamar Mitts (2025), Jacob Hale Russell & Dennis Patterson (2025), Michael Patrick Lynch (2025), Adam Szetela (2025), Matt Grossman & David A. Hopkins (2024), Daniel F. Stone (2023), Sigal R. Ben-Porath (2023), Sandro Galea (2023), Jacalyn Duffin (2022), Joel Simon & Robert Mahoney (2022), Chris Bail (2021), Siva Vaidhyanathan (2021), Linda Radzik (2020), Justin Tosi & Brandon Warmke (2020), Cailin O'Connor & James Owen Weatherall (2019), Morgan Marietta & David C. Barker (2019), and Jaime E. Settle (2019).

Stop Them, political scientist Barbara F. Walter explicitly links social media's incentives to social strife:

> It turns out that what people like most is fear over calm, falsehood over truth, outrage over empathy. People are far more apt to like posts that are incendiary than those that are not, creating an incentive for people to post provocative material in the hopes that it will go viral. With the introduction of the like button, individual Facebook users were suddenly being rewarded for posting outrageous, angry content whether it was true or not. Studies have since shown that information that keeps people engaged is exactly the type of information that leads them toward anger, resentment, and violence. When William J. Brady and his colleagues at NYU analyzed half a million tweets, they found that each moral or emotional word used led to a 20 percent increase in retweets. (2022, 110)

Social media content that is confrontational and aggressive is often effective at getting users to stop scrolling on their social media feeds, pay attention to a specific piece of content, and engage with it.[3] In many cases, this engagement involves trading insults with others. While such engagement may not be particularly constructive, it can help social media content receive a boost from platforms' curatorial algorithms, which of course increases its reach and guarantees that it will receive more attention than it would otherwise.[4] Online attacks are especially potent when they involve high-profile individuals such as entertainers and political leaders. Selecting famous individuals as targets of aggressive content helps broaden its appeal, as social media users who are not familiar with the person who posted the content may still have their interest piqued and choose

[3] Bail describes the connection between engagement and political extremism in the realm of social media: "research indicates that political extremists are pushed and pulled toward increasingly radical positions by the likes, new follows, and other types of engagement they receive for doing so – or because they fear retribution for showing any sympathy toward the mainstream. These types of behaviour mirror the famous finding of the social psychologist Leon Festinger about a doomsday cult from the 1950s: the further people become committed to radical views, the more difficult these commitments become to undo, and the more people come to rely on the status and support system that cults create." (2021, 66) It is important to note that evidence indicates that social media users are motivated not only by the desire to attract new contacts, but also the fear of losing existing ones.

[4] Social psychologist Jonathan Haidt expresses alarm at the volume of the antagonistic content that has flourished due to social media: "This new game encouraged dishonesty and mob dynamics: Users were guided not just by their true preferences but by their past experiences of reward and punishment, and their prediction of how others would react to each new action …The newly tweaked platforms were almost perfectly designed to bring out our most moralistic and least reflective selves. The volume of outrage was shocking." (2022)

to engage with it because they have an interest in the person being discussed. Users can accordingly capitalize on others' fame and followings in order to give their own content an advantage in the competitive world of social media.[5]

Philosopher Michael Patrick Lynch provides a useful portrait of social media communications that helps to convey the competitive dynamics that permeate online discussion. He specifically highlights the fact that social media users can be pressured into directing vitriol at others for the sake of winning the approval of onlookers and shoring up their own social position. Moreover, he points out that this kind of communication has become so pervasive that it has generated a lucrative industry across the globe that caters to those who wish to use social media to win political contests:

> It is not difficult to discern those aspects of contemporary culture that encourage performative political discourse. A chief one consumes our waking moments. Social media is an expressive machine—straightforwardly and consciously designed to encourage quick, nonreflective emotional engagement. We 'Like' posts, clicking on 'heart' symbols and emoticons symbolizing basic emotional reactions. We 'win' X or Facebook by hating on our enemies and conforming to the wishes of our cohort. We share posts without reading them because we know they make us seem informed or part of the team. Such behavior is ubiquitous and nearly universal … it is a major part of contemporary political discourse, employing political experts and consultants the world over, with easily shareable, expressible content encouraged and produced by billions of campaign advertising dollars. (2025, 44)

It would of course be wrong to suggest that only content that is divisive performs well on social media in terms of engagement. There are many types of content that go viral online, and a great deal of the time this has nothing to do with the promotion of outrage or anything of the sort.

[5] Renee DiResta highlights the competitive dynamics of social media: "Influencers compete for their audiences' time, which means that many hop from topic to topic. They generally maintain the same tone … This competition, however, leads many to present their opinions in increasingly extreme ways—they have to, in order to grab attention from both algorithms and human followers that reward moral righteousness, provocative claims, and outrageous rhetoric." (2024, 97)

Moreover, social media platforms are constantly tweaking their content moderation practices in order to compete with other platforms and retain users, so it would be an error to think that companies have a simple formula in place that enables them to keep users coming back to their platform over long periods of time. Social media content moderation practices have evolved dramatically since the 2000s, and it is virtually guaranteed that they will continue to evolve. Nonetheless, on balance, we can see that content that appeals to negative emotions such as anger enjoys a competitive advantage in the realm of social media insofar as it has proven itself to be a reliable means of capturing the attention of users and prompting them to interact with others online.[6] Social media companies are telling the truth when they tell consumers that their goal is to foster interaction among users, but unfortunately, in many cases these interactions are filled with anger and vitriol rather than good-faith communication.[7] As economist Daniel F. Stone puts it: "The fact that posts and tweets loudly expressing anger toward the out-party are more likely to go viral can incentivize strategic outrage and distortion for users trying (perhaps unconsciously) to maximize engagement, making (false) outrage-infused content even more common." (2023, 126)

i.ii: Social Media Incentives and Public Discourse

It turns out that incentivizing anger and vitriol has real implications with respect to the range of views that social media users encounter online. Empirical evidence indicates that we now

[6] Tobias Rose-Stockwell summarizes how these incentives operate: "The sheer quantity of content we're exposed to on a regular basis ensures that the type of content that we regularly like on social media will be emotionally charged … Discourse has always been polarized, but social media has amplified the ratio of extremely polarized content enormously. Through the dominance of these tools … we've watched our common discourse turn ugly, divisive, and increasingly polarizing. This is how small indiscretions can become massive cultural moments of moral judgment." (2023, 95)

[7] Tom Nichols states: "…social media is making us meaner, shorter-fused, and incapable of conducting discussions where anyone learns anything … Sometimes, human beings need to pause and to reflect, to give themselves time to absorb information and to digest it. Instead, the Internet is an arena in which people can react without thinking, and thus in turn they become invested in defending their gut reactions rather than accepting new information or admitting a mistake… (2024, 116-117)

find ourselves in a situation wherein ideologues often dominate political discourse on social media, while people with more moderate views generally refrain from participating in online debate. Philosophers Justin Tosi and Brandon Warmke note that social media discourse is now generally inhospitable to people with moderate views, instead favouring those with extreme views:

> …many people are being polarized to partisan extremes. Many of the moderates who remain in the middle, however, have had enough of their friends' contributions to public discourse. Indeed, those who are checking out of political discussion are disproportionately moderates. A recent study shows that, by and large, political extremists are the only people who devote much of their social media activity to discussion of politics. (2020, 89)

Sociologist Musa al-Gharbi shares a similar insight:

> Among Americans who use social media at all, the overwhelming majority (70 percent) rarely, if ever, post or share content about political or social issues … most commonly out of fear that they will be maligned or attacked for their views, or that their posts will otherwise be used against them … Research has found that the type of people who do use social media for political purposes tend to be very different from most others in terms of their dispositions both online and off. They are especially likely to be aggressive and status hungry. They tend to enjoy offending others but are also more easily offended themselves. (2024, 194)

It is clear that the views expressed via social media are far from being a microcosm of society more broadly.[8] Social media users who firmly align themselves with a specific ideological camp reap rewards as they affirm and reaffirm their commitment to promoting a specific agenda, while others with less intense partisan affiliations fear vilification and ostracism for falling afoul of orthodoxies that have been constructed in online spaces.[9] In some sense, modern social media

[8] J.P. Messina offers further support for this idea: "…recent research suggests that incivility is likely to be ramped up on the extremes of the political spectrum … Incivility appears to be leading … people to disengage from politics, to distrust the political process and their peers, and to feel that there is no place for their voices in the national conversation, not to adopt the views at the extreme ends of the spectrum. It appears that uncivil behavior leads moderates (the largest group of bystanders) and others to check out—not to march under any particular political banner." (2023, 49)

[9] Bail provides insight about confrontational behaviour on social media: "…extremists bond with each other by launching coordinated attacks on people with opposing political views. Though it may seem that social media extremists are most concerned with taking down the other side through superior argumentation … my research suggests that these attacks also serve a ritual function that pushes extremists closer together." (2021, 62)

platforms are the opposite of what consumers were promised when social media first began to gain mainstream traction. Instead of functioning as spaces wherein a vast array of individuals and groups from around the world can come together to discuss ideas[10] in a manner that is freewheeling and fluid, they function as spaces wherein rival ideological camps can attack their opponents and punish their own members for perceived disloyalty, thereby perpetuating rigidity.[11]

These patterns of behaviour put significant pressure on social media users to loudly proclaim views that are in alignment with a specific ideological camp, and to remain muted if and when they hold views that are disfavoured by this camp.[12] In other words, social media platforms encourage their users to conform. Again, the work of Lynch is helpful here. This researcher provides a helpful summary of how people make use of social media platforms in order to broadcast their allegiance with societal camps:

> For many people, their political commitments display their aspirational social identities … political posts signal whose team you are on. In sharing a meme or a 'news' article, people display their identities to others—thus reinforcing their actual or aspirational membership in their 'team' or tribe. And that goes for their commitments too: people often commit to propositions and ideas because they want to conform to, and wish to be seen as conforming to, a social identity. (2025, 47)

[10] The original mission statement for Facebook was "[t]o give people the power to share and make the world more open and connected". In 2017, the company revised its statement. The new version stated that the company's mission was "[t]o give people the power to build community and bring the world closer together." Interestingly, the latter apparently emphasizes unity and cooperation to a greater extent than the original.

[11] DiResta states: "Members of groups tend to reinforce each other's views, often moving each other toward a more extreme point than where they started. Factions appear to coalesce around the opinions of the most forceful members, and those who hold differing opinions—maybe more moderate—don't express them for fear of being ostracized. Since beliefs are shaped collectively, new information that conflicts with the group identity or comes from someone with an 'outside' identity can simply be rejected; this is one reason that partisans easily dismiss fact-checking if it comes from the 'other side.'" (2024, 127)

[12] Taylor N. Carlson and Jaime E. Settle argue that individuals' desire to enjoy group affiliation can outweigh their desire to make statements that are correct: "Individuals tend to feel good about themselves when they identify with and conform to groups that they value … By expressing the same opinion or providing the same answer as those in the group, they might be more likely to be included in the group; even if they are giving an incorrect answer to an objective question or an ill-informed opinion, at least the whole group will be wrong together." (2022, 25)

Social media's role in promoting rigid group loyalty is a key observation that animates this work. Social media platforms act as engines of conformity as various ideological camps use social pressure to ensure that their preferred orthodoxies are protected from scrutiny. Instead of functioning as spaces wherein users can entertain many ideas, take time to reflect on them, and then reach their own conclusions about which ideas are the most sound, social media platforms function as spaces wherein users face significant pressure to pick a side and remain firmly committed to it in order to remain in good standing with that side. While social media optimists may have once plausibly predicted that these platforms would have a liberating effect on discourse, and empower users to explore a broader palate of ideas than any previous technology had offered,[13] we now have reason to reach the conclusion that they are actually having a stifling effect on discourse. Political scientists Matt Grossmann and David A. Hopkins offer a clear summary of these dynamics in the United States context:

> …for many Americans, activists' rhetoric helps to communicate the appropriate attitudes maintained by their side. Punishing prominent people who violate the prevailing norms of the moment – such [as] the now common practice of social media shaming, a favorite tactic of ideological purists on both sides – can both promote these values and demonstrate that dissent will jeopardize one's standing in the social group. (2024, 70)

Over the course of the late 2010s and early 2020s, the phrase "cancel culture" entered the popular lexicon, and the topic has become so prevalent that it has been taken up by academic researchers.[14] It is now common for people in all sorts of professional domains to comment on

[13] Jack M. Balkin explains: "The early promise of social media, like the early promise of the internet generally, was that they would promote a diversity of views and offer alternatives to dominant cultural gatekeepers. They would also support the growth and spread of knowledge by lowering the costs of knowledge production, dissemination, and acquisition. To some extent, this promise has been realized. Widespread access to digital communications has also helped people scrutinize professions and institutions and disclose their flaws and failings. But social media, like the internet more generally, have also disrupted norms of civility, undermined professionalism, and helped people spread distrust in knowledge-producing institutions and in democracy itself." (2022, 242, in Bollinger and Stone, ed.)

[14] Sigal R. Ben-Porath offers the following summary of the cancel culture phenomenon: "The depiction of 'cancel culture' is commonly negative, portrayed as an exaggerated and even anti-democratic response to any small offense,

cancel culture, and many are concerned about the possibility of one day being a target of cancel culture themselves. While this concept eludes precise definition, as it can mean different things to different people, the phrase generally denotes a social dynamic wherein offensive conduct by individuals is given a spotlight in a manner that is uncharitable, punitive, and unforgiving. To cancel someone is to promote the idea that they are unworthy of being liked or listened to, and to create a stigma that can be attached to those who choose to collaborate with them or take an interest in their views. Philosopher Linda Radzik states: "A recent variation on naming and shaming involves declaring that the wrongdoer is 'canceled.' Such declarations seem to operate as both public shaming and calls for social withdrawal from the condemned person." She explains: "By declaring a wrongdoer canceled, one resolves, and encourages others to resolve, to deny the wrongdoer a public platform." (2020, 49) In many cases, cancellation involves explicit calls for a person to be fired by their employer. While cancellation is certainly not a form of exile in a literal sense, it can amount to a form of social exile whereby an individual's ties (especially professional ties) to other individuals and groups are severed.

While it would be an error to think that cancel culture can only take hold in the realm of social media, it is easy to see why this type of media can act as an exceptionally potent venue for cancellation efforts. Social media is inherently interactive, and users can track the impact of their posts in real time. While older media requires some significant amount of time to pass before a

sweeping up innocent individuals and companies in its wake. Under this view of 'cancel culture', a person making an insensitive remark with no harmful intent and an institution or a company failing to abide quickly enough by an ever more intricate public demand for adherence to the ideological orthodoxy of the day are subject to 'cancellation': a public outcry calling for firing the offending individual, boycotting the company, or 'abolishing' the institution." (2023, 68-69) Matthes explores the issue of cancel culture in the context of the art community: "As I understand it, cancel culture in the arts is characterized by widespread dispositions to engage in automatic calls to boycott and ostracize artists based on their immoral actions or words ... Cancel culture may often operate in a way that begins with call-outs as a tool, but cancel culture in the arts ultimately aims to erase rather than excoriate." (2021, 78-79)

publisher can learn whether the content they have circulated has succeeded in gaining traction, social media platforms provide a potentially endless supply of instant feedback.[15] If a cancellation effort gains traction on social media, users will find this out immediately, and can accordingly double down on their efforts. The highly gamified[16] environment of social media easily lends itself to the notion that one is making a difference by sharing content that drives engagement. If one sees that their posts are attracting likes, shares, and follows, then this can generate feelings of empowerment among users. Radzik offers analysis that is instructive:

> Social media users enjoy a boost of dopamine when their contributions are liked and reposted. We receive psychological benefits from feeling that we are 'in the know' or part of a movement. We like to feel virtuous, to feel more virtuous than other people, and to have our virtue witnessed. People sometimes also join in naming and shaming campaigns for fear that silence will signal support for the wrongful action. They join in the shaming for fear of being shamed themselves. Add to this our susceptibility to the pleasures of vengeance and schadenfreude, and the extra temptation when we can indulge our aggressive impulses anonymously. All of these factors help contribute to the snowballing effect in online naming and shaming campaigns … (2020, 54-55)

Sociologist Chris Bail offers the following reflection regarding his own experiences researching hardcore partisans on social media:

> The symbolic meaning of the bonds that extremists make with each other became even more apparent to me when I learned how closely extremists monitor their followers. Though social media sites do not alert users when people stop following them, several of the extremists we interviewed used third-party apps to identify such individuals. People who unfollowed the extremists we studied – particularly several of the conservative extremists – were often subject to even more aggressive attacks than the moderates … For

[15] DiResta explains how abundant feedback can shape the words and behaviours of people who have amassed significant online followings: "Influencers are acutely aware of their engagement metrics, because creating content is their livelihood. It is a struggle, sometimes, to remain true to their vision … This is the realm of audience capture—a feedback loop in which creators produce content their audiences will approve of and gradually begin to internalize it themselves … That feedback loop … can drive a content creator into a particular niche that's difficult to escape, taking them in a completely inauthentic or extreme direction." (2024, 99)

[16] The Oxford Dictionary of English defines gamification as "the application of typical elements of game playing (e.g. point scoring, competition with others, rules of play) to other areas of activity, typically as an online marketing technique to encourage engagement with a product or service". The fact that social media users are given feedback about the popularity of not just their own posts, but also the posts of other users, indicates that the social media environment is gamified, and users are tacitly encouraged to compete with one another to ascend the social media ranks.

> me, this type of retribution further underscores how deeply trolls value the status and influence they achieve online, and how much it upsets them when people on their own side sever ties with them. (2021, 65)

This research indicates that the desire for attention and approval shapes a great amount of online discourse, fuelling extremity and antagonism. Moreover, social media is noteworthy for its propensity to foster interaction between low-profile individuals and high-profile individuals, and some users may be attracted to the prospect of inflicting significant reputational and professional damage on people who outrank them in terms of popularity, influence, and wealth. There is often something exciting about an underdog defeating a more powerful opponent, and the dynamics of social media make it remarkably easy for users to conceptualize themselves in this role.[17]

i.iii: The Distinction Between Cancellation and Intimidation

Although cancel culture is now a popular topic of discussion, it is not the case that a consensus has been reached about whether it is truly worthy of significant alarm and resistance.[18] Debate is still unfolding as to whether cancel culture is a genuine phenomenon, or whether it is really just an unflattering label for a normal societal process wherein people face accountability for their words and actions after they have hurt others in some way. A skeptic could point to the massive commercial success of various reactionary media personalities as evidence that cancel culture is a myth, as these individuals routinely comment on controversial issues, and offend significant portions of the general public, without facing serious personal or professional

[17] Francis Fukuyama notes that the incentive structure of social media can generate feelings of importance: "Social media companies have cleverly created incentive systems that persuade people they are doing something important if they pile up 'likes' or retweets, whereas in reality such measures are significant only within the closed environment of social media itself. This is not to say that social media cannot lead to meliorative outcomes in the real world. Most people, however, are satisfied with the simulacrum of reality that they get through their online interactions." (2022, 112-113)

[18] Adrian Daub, a vocal critic of those who raise alarm about cancel culture, states: "Cancel culture anecdotes are tendentiously composed fables, often based on only one source, which are—at least in the United States—usually purveyed and promoted by politically motivated actors." (2024, 20)

repercussions. Indeed, there is room for reasonable debate about the prevalence of cancel culture, what exactly cancel culture entails, and whether our modern era is really so different from past eras with respect to the issue of public accountability for offensive actions.

Debates about cancel culture can be illuminating, as they can help us develop a deeper understanding of the period in history that we happen to occupy, and the extent to which it parallels other moments in history.[19] However, such debates are separate from the core concern that animates this dissertation. Accordingly, a clarification is warranted in order to prevent misunderstanding. The concern about social media that animates this discussion is not that this technology is functioning as some kind of cancellation machine and causing wide swaths of the population to suffer personal and professional ruin. It could be that only a small portion of the overall population is having such experiences. While it may not be the case that large segments of the population are undergoing cancellation episodes as a result of social media controversy, it is clear that in the social media age, very many people are reluctant to express their genuine views out of fear of social punishment.[20] In other words, while it may not be the case that cancellation is rampant in the social media age,[21] it is the case that intimidation is rampant. Self-censorship has

[19] Jonathan Rauch argues that there are parallels between contemporary cancellation campaigns and the ostracism of gay people throughout recent decades: "We gay people are very, very well acquainted with canceling … We did not spend the last half century and more fighting against it so that we could turn the tables and make pariahs of others." (2021, 254-255)

[20] James L Gibson and Joseph L. Sutherland offer the following insights about the prevalence of self-censorship in the United States: "Over the period from the heyday of McCarthyism to the present, the percentage of the American people not feeling free to express their views has tripled. In 2020, more than four in ten people engaged in self-censorship. Our analyses of over-time and cross-sectional variability suggest that, first, self-censorship is connected to affective polarization among the mass public, with greater polarization associated with more self-censorship … micro-environment sentiments, such as worrying that expressing unpopular views will isolate and alienate people from their friends, family, and neighbors, may be the driver of self-censorship." (2023, 1) While my discussion operates on the premise that social media plays an important role in intimidation and the self-censorship that comes with it, I do not claim that social media is the sole factor involved in generating this phenomenon.

[21] Another way of phrasing this is to say that while cancellation may be common in the modern era, there is no reason to think that it is more common now than it was in previous eras. One might argue that people have always been punitive, and that over time they have simply chosen to be punitive in different ways. For example: while it is now

increased dramatically over the course of the 2010s and 2020s, and this development may have a pernicious influence on society that merits thorough philosophical investigation.

The core concern about social media that animates this discussion is that this technology is facilitating fear of social censure, thereby helping to generate immense chilling effects,[22] and that these chilling effects are making society worse off by undermining important social goods. If it is true, as empirical evidence suggests, that online intimidation is spilling over into offline facets of life in addition to the online domain, then this gives us further reason to worry about the ability of online discourse to pressure institutions and individuals to behave in ways that are detrimental to themselves and society more broadly. Rather than making a case against cancel culture per se, this dissertation aims to make a case against what we can reasonably call "intimidation culture". Significant portions of the forthcoming discussion will be dedicated to examining reasons why intimidation culture ought to be understood not merely as an irritant, but as a force that has the ability to degrade society in various ways, and make life less fulfilling for the individuals who inhabit said society.

i.iv: Weighing Costs and Benefits

It is of course necessary to note that for some, the ability of social media platforms to highlight the offensive words and deeds of various individuals and institutions is something that

relatively common for individuals to be fired from their jobs after having made homophobic remarks, in the past it was common for individuals to be fired after having been outed as homosexual. (See Rauch 2021, Chapter 8) In both cases, punitive action is present, but the former case is congruent with contemporary social norms, while the latter is not. The upshot is that one can accept the view that cancel culture is prevalent today while rejecting narratives about its rising prevalence on the grounds that it has always been prevalent in one form or another.

[22] Alycia Burnett, Devin Knighton, and Christopher Wilson describe a "silent majority" in their discussion of self-censorship dynamics on social media. Their research indicates, in line with other findings, that "the hardcore vocal minority opinion will not be silenced … and they will speak up about the issues they care about". (2022, 7) They present evidence that small groups of hardcore ideologues are able to chill discourse on social media.

ought to be welcomed on the grounds that it can generate accountability and lead to meaningful societal gains. Rather than viewing antagonistic social media discourse as an engine of intimidation, certain observers may view it as a conduit for forms of activism wherein unsavoury components of society are held to the proverbial light so that the public may extract valuable lessons from them.[23] Some might allege that the portrait of social media discourse that has been offered in this chapter is too pessimistic, as it does not adequately reflect the extent to which social media platforms have aided people in raising awareness about important causes and demanding changes to the status quo. Indeed, it would be wrong to deny that in some cases, this is precisely what is achieved via social media discourse wherein particular persons and groups are given a spotlight for their bad acts. It is sometimes the case that deploying the immense power of the Internet in order to draw attention to various misdeeds can help yield needed reforms and improved behaviour, as public pressure is a remarkably effective tool for producing these kinds of changes.

While it is true that the dynamics of social media can empower users to demand repercussions for various forms of misbehaviour, many have noted that the maximalist tendencies of online platforms can often lead to disproportionate reactions towards those who find themselves being targeted with online castigation. Philosopher Erich Hatala Matthes, who pays particular attention to the impact of cancel culture in the domain of the arts, offers the following analysis:

> We can endorse the position that people should be held accountable for what they say without thinking that scorched earth is always the appropriate response. There's a significant difference between a bigot raining slurs and a thoughtful person making a good-faith effort to articulate a view that you think is ultimately wrong, or even harmful: cancel

[23] Siva Vaidhyanathan explains how social media discourse can shape activist movements: "Social media services … do affect political and social movements, and thus the protests that ensue, in particular ways. The presence of Facebook does not make protests possible, more likely, or larger. But Facebook does make it easy to alert many people who have declared a shared interest in information and plans. It lowers the transaction costs for early organization. Most important, Facebook has the ability to convince—perhaps fool—those who are motivated and concerned that …'we are not alone.'" (2021, 132-133)

> culture often seems to erase the differences between the responses that these cases call for, so it's not hard to understand why criticizing it has become fashionable … (2022, 80-81)

Matthes argues that the destructiveness of cancel culture "is perhaps best displayed in contexts where it targets everyday people rather than famous artists." He explains:

> The canonical case has become the story of Justine Sacco, who after tweeting a tasteless joke found herself the target of an internet outrage mob and ultimately fired from her job. Similar instances abound. These cases illustrate cancel culture's inability to operate at an intensity that isn't turned up to 11, which isn't really surprising: you can't partially erase someone—it's all or nothing. Its advocates may want accountability, but accountability should be modulated by a commitment to proportionality. There's considerable space between holding someone accountable for a bad joke and trying to get them fired. (2022, 80-81)

Even if we accept the notion that, in principle, social media discourse can function as a pathway towards fairness and accountability, it is incumbent upon us to examine and reflect upon the ways in which attempts to call people to account via this technology can misfire and produce significant injury to persons and communities, while leaving larger issues unaddressed.[24] Any sensible discussion of social media and its societal impacts will need to consider the potential drawbacks associated with online castigation in addition to its potential benefits. While this dissertation will not advance the extreme and rigid position that the power of social media should never be used to direct harsh criticism towards individuals and institutions that have engaged in damaging conduct, it will offer an emphatic case that the societal costs associated with vitriolic

[24] Sarita Srivastava argues that in many cases, the spectacle of "calling out" serves to make participants feel gratified, despite the fact that little is actually being done to address longstanding injustices: "The social media spotlight, or 'calling out,' of … individuals also help many people feel vindicated. They may feel that justice is being done, and that racist people are getting justly punished … However, there is also a corollary: if they are guilty, then I am innocent. Everyone who has escaped being 'called out' also feels more virtuous … the incident and the person become spectacles viewed from afar, with little connection to the long history and ubiquity of anti-Black racism. Social media callouts also further cultivate superficial declarations of antiracism as a moral position, rather than as a practice." (2024, 232-233)

and accusatory online discourse are great, and that it is sound to seek changes to our media ecosystem so that these societal costs can be reined in.

This is a far cry from arguing that the project of holding people to account should be abandoned. Rather, it simply indicates that people who wish to achieve accountability ought to be cognizant of the fact that the dynamics of online discourse can produce impacts that go well beyond the domain of accountability, and ought to question whether participation in spectacles of online shaming is the most appropriate strategy for realizing this objective. As venues that reward extreme and attention-grabbing activity, social media platforms can facilitate behaviour that is downright destructive, and this has the potential to impose costs on society at large in addition to particular individuals. In order to better appreciate how the dynamics of social media can cause efforts to achieve accountability to spiral into something much less noble and much more ugly, let us consider the words of communications scholar Jason Hannan:

> As with public punishment in the premodern world, punishment on social media has three core features. First, online punishment must produce pain and torment in the form of shame. Second, online shaming takes a ritual form. As with physical torture, online shaming marks the accused with the truth of their guilt. Third, online shaming is a public spectacle, a virtual theatre of cruelty, in which witnesses are encouraged to laugh, jeer, hector, and abuse the guilty. (2023, 76)

Hannan goes on to state: "What is termed 'call-out culture' is driven by the sadistic pleasure of witnessing public spectacles of shame and humiliation. If the accused is guilty, then witnesses need not feel guilty about watching them suffer." (2023, 77)

Hannan's evocative commentary is helpful for understanding how the dynamics of social media discourse can lead to unsettling forms of behaviour that go far beyond any reasonable pursuit of accountability. While it is indeed possible for online criticisms to be measured and

rehabilitative in character rather than extreme and punitive, the latter type of content stands a better chance of being amplified in the realm of social media in many cases.[25] The architecture and incentives of social media platforms make them highly conducive to forms of communication that are overzealous and designed to attract attention, and this is something that must be borne in mind by those that view these technological tools as assets to the project of achieving accountability throughout society. It would certainly be an overstatement to suggest that it is impossible for people to use these platforms responsibly in pursuit of reasonable changes to the status quo, but it is not an overstatement to point out that antagonistic social media discourse often tends to escalate in terms of intensity and toxicity, thereby generating reactions to perceived wrongdoing that are disproportionate, and perhaps even deliberately cruel in nature. In many cases, the costs associated with online castigation outweigh the benefits, and this costliness is a theme that will inform much of the discussion that will be offered over the course of the coming chapters.

The chapter that follows will turn our focus away from the nuances of social media communication and towards the philosophy of free expression. A normative vision of free expression will be outlined that involves much more than just the absence of heavy-handed state intervention into matters involving expressive acts. Instead, this chapter will advocate for a social and political system wherein proactive efforts are made to facilitate good-faith dialogue between individuals and groups with diverse worldviews. It will be argued that when an atmosphere conducive to such dialogue is created and sustained, this provides opportunity for a set of key social goods to flourish, thereby generating significant gains for communities over the long term.

[25] Mary Beth Willard states: "Social media does not encourage careful deliberation or taking time to think before posting or reacting, and so it is capricious. More worryingly, social media rewards moral grandstanding, publicly posturing that one is on the right side of the issues by posting or reacting in the right way … participating in canceling risks undermining genuine ethical discussion and genuine ethical growth." (2021, 26)

The specific social goods that will be brought to the fore are critical intellectual faculties, authenticity in discourse, and equity in accountability. If we are interested in the project of maximizing utility as well as the long-term sustainability of society, then these social goods ought to be appreciated and actively cultivated.

Next, in Chapter 3, it will be argued that the dynamics of contemporary online discourse are at odds with this normative vision of free expression, and that online intimidation culture is injurious to the social goods that free expression helps to secure. A case will be made that online intimidation culture undermines the ability of societies to deploy their intellectual capital appropriately, thereby making them more vulnerable to an array of internal and external threats. Chapters 4 and 5 will examine and assess a variety of potential remedies to the problem of online intimidation culture, and strive to identify which are most aligned with the normative philosophical principles set forth in Chapter 2. Chapter 6 will make the case that by embracing a substantive commitment to free expression and resisting the pressures of intimidation culture, institutions can bolster their credibility and cultivate social trust, which is vital for avoiding the deepening of divisions throughout society. The seventh and final chapter will explore more optimistic philosophical territory, arguing that despite its many dysfunctions and shortcomings, the technology of social media still has the ability to be beneficial to free expression and intellectual diversity. This concluding chapter will explore how social media can help cultivate understanding, compassion, and cooperation between diverse individuals and groups across the globe, thereby augmenting the wellbeing of human populations and cultivating a more peaceful and prosperous world.

Chapter ii: A Millian Atmosphere of Free Expression

ii.i: Free Expression in a Millian Framework

The goal of this chapter is to outline a normative vision of free expression that will inform the commentary about social media that is offered in later portions of this dissertation. While we will eventually confront the issue of whether and how the dynamics of contemporary online discourse can be injurious to free expression, our present task is to develop a better understanding of what exactly free expression entails and how it can produce societal benefits. Does free expression merely involve the absence of aggressive state policies that seek to constrain expressive acts, or should free expression be understood in a broader manner? How exactly does free expression generate gains for society that are worth caring about and working to preserve? In order to answer these questions, I will invoke the philosophy of the seminal liberal thinker John Stuart Mill. As we will see, Mill's writings can be helpful not only for constructing a negative case against various forms of censorship, but also for constructing a positive case in favour of a broad array of societal actors making efforts to cultivate an atmosphere wherein intellectual diversity and good-faith deliberation can thrive.[26] The normative vision of free expression that is advanced in this chapter is accordingly more complex and demanding than alternatives wherein free expression is conceptualized merely as a system of protection from various forms of aggression in response to expressive acts.

Mill's 1859 text *On Liberty* offers what is widely considered to be the deepest and most well-developed account of the importance of free expression that has been published to date. In it, Mill argues that "the appropriate region of human liberty" involves not only "absolute freedom" with respect to one's thoughts and opinions, but also the opportunity to share these thoughts and

[26] The language of negativity and positivity invoked here aligns with the conceptual distinction made by Isaiah Berlin in his 1958 essay "Two Concepts of Liberty". (Berlin 2002)

opinions with others.[27] Intellectual freedom and expressive freedom are so intimately intertwined for Mill that he views them as being "practically inseparable". (2015, 15) Accordingly, Mill provides a series of powerful arguments in favour of limits being placed on what states, including those that enjoy popular support, can do with respect to policing the expressive acts of individuals and groups throughout society.[28] He emphasizes that censorship of this kind is costly not only to those who are its direct targets, but to society as a whole, as it deprives people of the opportunity to grapple with the prohibited speech in question and accordingly arrive at a clearer understanding of the truth. Mill compellingly argues that even errant views can have edifying effects as they prompt people to carefully consider their reasons for embracing certain positions over others, rather than simply accepting specific views in an unreflective manner. He warns that efforts to purge discussion of unpopular or noxious views from society can have stultifying effects that undermine the intellectual health of a populace. While Mill's discussion of the harm principle[29] does provide philosophical grounds for states to legitimately intervene in matters involving expression when expressive acts directly jeopardize the safety and wellbeing of their targets,[30] he generally maintains that states must be required to abstain from meddling with matters of

[27] Mill explains: "This, then, is the appropriate region of human liberty. It comprises, first, the inward domain of consciousness; demanding liberty of conscience in the most comprehensive sense; liberty of thought and feeling; absolute freedom of opinion and sentiment on all subjects … The liberty of expressing and publishing opinions may seem to fall under a different principle, since it belongs to that part of the conduct of an individual which concerns other people; but, being almost of as much importance as the liberty of thought itself, and resting in great part on the same reasons, is practically inseparable from it." (2015, 15)

[28] Mill explicitly argues that popular support cannot legitimize government action that is excessively coercive: "Let us suppose … that the government is entirely at one with the people, and never thinks of exerting any power of coercion unless in agreement with what it conceives to be their voice. But I deny the right of the people to exercise such coercion, either by themselves or by their government. The power itself is illegitimate." (2015, 19)

[29] Mill famously states: "the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise." (2015, 13)

[30] For a thorough discussion of Mill's harm principle, see Chapter 2 of L.W. Sumner's *The Hateful and the Obscene* (Sumner 2004).

expression, including those that are highly controversial, so that individuals and groups can explore ideas in a freewheeling fashion and determine for themselves which ideas and ways of living are the most sound. Limits on state action are essential in order for intellectual freedom and expressive freedom to thrive, along with many other forms of freedom with which they are connected.

In order to refine our understanding of Mill's approach to delineating the scope of legitimate state intervention into matters involving expression, we can look to the writings of philosopher L.W. Sumner. Sumner notes that while the harmfulness of expressive acts is indeed one necessary precondition for legitimate state interference with such acts, more criteria need to be met in order for constraints on expression to be legitimate within a Millian framework. Sumner provides the following explanation of the conditions that must be satisfied in order for specific limits on expression to be justified:

> If … restrictive legislation manages to pass the harm test it does not follow, however, that it is justified by Mill's liberty principle. That principle makes harm to others a necessary condition for limiting liberty, but not a sufficient one. The legislation must also pass a cost-benefit test: restricting the expression in question must yield a better balance of benefits over costs than leaving it unregulated. This requirement of a positive cost-benefit balance does not provide a simple algorithm for deciding whether, and when, the state is entitled to enforce restrictions on forms of expression in those cases in which the harm test has been satisfied. However, it does suggest the kinds of factors which will be relevant. First, the restriction must have some reasonable expectation of success. While it may be thought desirable to inhibit or suppress some form of expression by legal means, it is a further question whether doing so is possible. To the extent that the restrictions can be readily circumvented, by an underground market or by technological innovations such as the Internet, the case for them is weakened. Second, there must be no less costly policy available for securing the same results. Even when it promises to be effective in preventing some significant social harm, censorship abridges personal liberty and deprives consumers of whatever benefits they may derive from the prohibited forms of expression. It should therefore be the last, not the first, resort of government for preventing the harm in question. Where less coercive measures (education, counterspeech, etc.) promise similar results they should be preferred. Where a narrower infringement of freedom of expression will be equally effective it too should be preferred. Third, the expected benefits of the restriction must, on balance, justify its costs. Censorship can compromise other important social values, such as vigorous engagement in public debate. It can have a 'chilling effect' on

> legitimate forms of expression (literary, artistic, etc.). However well intended the restriction might be, in practice it will be administered by police, prosecutors, judges, or bureaucrats who may use it to justify targeting unpopular, marginal forms of literature with no significant capacity for social harm. On balance, the benefits to be gained by legal restraints on expression must be great enough to justify the collateral costs. (2009, 207)

Sumner does an excellent job of outlining the balancing process that must take place in order to assess whether limitations on expression are legitimate from a Millian perspective. In order for limits on harmful speech to be justified, these limits must have a real likelihood of achieving their objective, they must be the least costly means available of realizing this objective, and the societal gains of enforcing the limitations in question must be greater than the costs that are associated with them. It follows from this set of criteria that when considering whether a particular form of expression is harmful, we must also consider whether calling upon the state to censor this expression may produce its own set of harms. This is why reasonable people with liberal orientations can reach different conclusions about whether censorship is appropriate in an array of particular cases. For instance, Sumner carefully examines the issue of hate speech, and while he does not dismiss the notion that hate speech is likely to inflict harm on individuals and groups, he calls into question whether using the machinery of the state to punish such expression is the wisest course of action for those who wish to alleviate these harms. He points out, compellingly, that other mechanisms such as counterspeech are available that can mitigate these harms while simultaneously leaving the goods associated with free expression intact, and thereby imposing fewer costs on society.[31]

[31] Sumner offers commentary about a case involving David Ahenakew, a Canadian leader who made a series of hateful comments: "… criminal sanctions should be employed only 'when the harm caused or threatened is serious, and when the other, less coercive or less intrusive means do not work or are inappropriate'. Both discrimination and hate violence certainly qualify as serious harms. However, there appear to be less intrusive means available of neutralizing any contribution that hate speech might make to these practices. One of these means is precisely the kind of counterspeech elicited by Ahenakew's remarks. … The antiracism cause was arguably better served by having Ahenakew speak his mind and arouse a firestorm of opposition than it would have been had he been intimidated into silence by the fear of prosecution." (2004, 194-195)

My contention is that similar reasoning ought to be deployed in discussions of social, rather than political, constraints on expression. It is surely the case that a wide range of ideas and opinions may justifiably be viewed as wrong or immoral. Every day, people across the globe give voice to attitudes that are pernicious and deserving of serious resistance. In many cases, the people who are exposed to this expressive content may consequently experience stress or pain, which can reasonably be conceptualized as forms of harm. However, if we are going to deploy a Millian framework when exploring these issues, then our investigation cannot end simply with the observation that a particular form of expression generates harm. An additional question that must be confronted is whether targeting these speakers with social punishment is the best strategy for dealing with their controversial expressive acts. A Millian approach to these issues requires us to countenance the possibility that levying social sanctions upon such speakers, thereby causing them to experience exclusion and ostracism, might actually make society worse off in the aggregate. Even if the instinct to rebuke the words and ideas of such individuals is entirely warranted, we must not lose sight of the possibility that social punishment is, on balance, a poor strategy for addressing the harmfulness of their ideas and speech. We must ask whether the costs of social punishment outweigh the benefits, and also whether alternative strategies for dealing with the expression are available that are less costly at the societal level.

A key component of Mill's philosophy of free expression, and one that is particularly relevant for our purposes, is his view that while protections from gratuitous state intervention are extremely important, they are also insufficient, as social tyranny can be just as stifling to individuals and groups as political tyranny. Let us consider the following statement from Mill:

> Protection ... against the tyranny of the magistrate is not enough: there needs protection also against the tyranny of the prevailing opinion and feeling; against the tendency of

> society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them; to fetter the development, and, if possible, prevent the formation, of any individuality not in harmony with its ways, and compels all characters to fashion themselves upon the model of its own. (2015, 8)

This statement emphasizes that while free expression can indeed be threatened by formal state censorship, it can also be undermined by actors outside of the state.

Radzik, who invokes Mill in her own discussion of social punishment, states:

> The philosophical literature on punishment is so wholly concentrated on the state's responses to crime that authors sometimes dismiss talk of punishment in everyday life as merely metaphorical. But this is mistaken. Legal norms are not the only ones that society enforces, and the mechanisms of law are not the only methods of enforcement that society uses. (2020, xii)

This observation is instructive for our purposes. We can read Mill as advancing the normative view that while it is important for members of society to be protected from government overreach with respect to matters of expression, it is also important for them to be protected from overreach by actors that are separate from the government.[32] In order for members of society to enjoy a substantive "atmosphere of freedom" in addition to a formal guarantee of free expression, non-state actors must behave in a manner that gives individuals room to explore ideas in an uncontrived fashion,[33] rather than using social rewards and punishments to pressure them into embracing a certain set of beliefs.

---

[32] Cass R. Sunstein states: "*On Liberty* is widely taken to be an argument for limited government, and so it is. But it is crucial to see that in contending that people may be restrained only to prevent 'harm to others,' Mill was calling for restrictions on social norms and conventions, not merely on government. Much of his attack was aimed at the oppressive quality of public opinion." (2025, 48)

[33] The cultivation of individuality is an important topic for Mill, and informs much of his discussion of ethics and politics. Accordingly, it is reasonable to ask why individuality has not been listed as one of the social goods that are presently threatened by online intimidation culture. The reason is that in a Millian framework, critical thinking is so tightly bound with the issue of individuality that it would be superfluous to offer discrete discussions of the two. When a person refines their critical intellectual faculties, their capacity for independent thought and conduct increases accordingly. While it is possible to conceptualize individuality and critical intellectual faculties as two separate social goods, Mill clearly views them as being inextricably linked:"… to conform to custom, merely *as* custom, does not educate or develop in him any of the qualities which are the distinctive endowment of a human being. The human faculties of perception, judgment, discriminative feeling, mental activity, and even moral preference, are exercised

Mill's reference to an "atmosphere of freedom" in *On Liberty* is noteworthy, and it stands in stark contrast with his reference to an "atmosphere of mental slavery" elsewhere in the text.[34] This language suggests that it is possible for unwritten and unspoken rules to interfere with human liberty, and stifle the intellectual contributions of people who are capable of offering more to the world given appropriate background conditions. Clearly, Mill thinks that human thought and behaviour is shaped by much more than the official policies that are enshrined by states, and this is why much of the discussion of *On Liberty* is dedicated to explicating and critiquing the many ways in which society can undermine individuality without ever deploying the machinery of government. Protecting human beings from figurative slavery and nurturing their freedom requires vigilance with respect to these pernicious social forces. Indeed, this text finds Mill dedicating far more of his attention and energy to cautioning his audience about the dangers of society's ability to promote conformity and stunt individuality in insidious ways than to the project of delineating the appropriate scope of state power.

It must be noted that enslavement of any kind is not only destructive to the freedom of human beings; it is also devastating to their ability to realize an identity for themselves of which they are the author. For Mill, freedom is inextricably linked with individuality, and the two cannot exist apart from one another. He views individuality as a sort of safeguard that shields societies

---

only in making a choice. He who does anything because it is the custom makes no choice. He gains no practice either in discerning or in desiring what is best. The mental and moral, like the muscular powers, are improved only by being used … If the grounds of an opinion are not conclusive to the person's own reason, his reason cannot be strengthened, but is likely to be weakened, by his adopting it …" (2015, 57-58) Here we see Mill arguing that there is a strong connection between the development of individuality and the strengthening of the human intellect. The two progress in tandem, and are undermined by conformity with prevailing ideas and behaviours.

[34] Mill highlights the importance of a society's atmosphere for good and for ill. He states: "There have been, and may again be, great individual thinkers in a general atmosphere of mental slavery. But there never has been, nor ever will be, in that atmosphere an intellectually active people." (2015, 34) Later, he makes the claim that: "Persons of genius … are, and are always likely to be, a small minority; but in order to have them, it is necessary to preserve the soil in which they grow. Genius can only breathe freely in an *atmosphere* of freedom." (2015, 63-64)

from an array of ills, and can be eroded in a variety of ways: he states that "[e]ven despotism does not produce its worst effects, so long as individuality exists under it; and whatever crushes individuality is despotism, by whatever name it may be called…". (2015, 62) This indicates that within a Millian framework, it is entirely possible for humans to inhabit communities and institutions wherein formal limits on expression are minimal or nonexistent, and yet still live in an atmosphere that is oppressive and denies them the opportunity to think and communicate with others in a manner that is conducive to the development of their individuality. The pressures of conformity can significantly undermine people's freedom and their individuality even when there are no codified rules in place requiring individuals and groups to conform. This means that in order for free expression to be adequately realized in a society, the society in question must provide an atmosphere wherein conformist pressures are reined in, in addition to an arrangement wherein government offices are constrained from prosecuting expression that they dislike.[35]

The importance of an atmosphere of free expression becomes clearer when we consider Mill's suggestion that "social tyranny" can be even more stifling than political tyranny on the grounds that "it leaves fewer means of escape", and can "[penetrate] much more deeply into the details of life". (2015, 8) It is not too difficult to imagine why Mill might think this. While members of a society might be successful at skirting formal rules regarding expression that are implemented by the state, as many states lack the ability to surveil their populations at all times and to prosecute every act that falls afoul of their laws, it is often more difficult for members of a society to succeed

[35] Michael J. Glennon argues that the founders of the United States had a robust understanding of the dangers associated with groupthink and conformity: "The Founders didn't frame their insights in modern terms of cognitive dissonance, pre-deliberation bias, and so forth, but they were familiar with what we now refer to as groupthink. They understood that orthodoxy can be suffocating, in insular government groups as elsewhere. They knew that iconoclasts, naysayers, dissenters, and boat-rockers were necessary to keep the nation's political and intellectual life vibrant. Crackling disagreement within each sphere, they believed, would be invigorating." (2024, 36)

at evading the judgment of the public when their words or conduct are at odds with popular social norms.[36] Philosopher J.P. Messina explains:

> … as John Stuart Mill reminds us, when mobs can too easily enforce their conception of what counts as acceptable speech, the result can be worse than state censorship. This is an important point: To the degree that it becomes easy to pressure powerful institutions to part ways with people who say things that press the boundaries of what is acceptable, the result can be an atmosphere in which no one feels very free to question social orthodoxies. It isn't hard to see why such a state of affairs is worrisome. (2023, 82)

Unless one wishes to live in complete isolation, which is an impractical project for most, they must have the ability to interact and cooperate with others to a certain extent. If a person finds themself on the receiving end of harsh judgments from other members of society, and is accordingly excluded from a broad array of social spaces in which they wish to be active, there is generally no clear process through which they can seek reconciliation, as ostracism is a fate that is brought about through mechanisms that are mostly (if not entirely) informal and voluntary. The nature of such exclusion means that there is often no authority that can be appealed to in order to determine whether exclusion is legitimate and appropriate in any given case.[37] Informal social processes can be even more potent than formal state processes when it comes to undermining free expression and creating an "atmosphere of mental slavery". Radzik raises concerns about the unconstrained nature of social punishment that are relevant for our purposes:

> Issues of proportionality … arise in legal punishment, of course, but they threaten to be more intractable for informal social punishment. Legal penalties are measured and doled out by a central authority, but public shaming is uncontrollable. The original namer cannot determine how many people will eventually be included in the audience, what their evaluative reactions will be, or what they will do with the information in the future. The original punisher may express his censure in measured and morally nuanced language, yet set off a firestorm of indignation … that includes loss of employment, hate speech, or

[36] Mill notes that social norms can be powerful engines of obedience: "Wherever the sentiment of the majority is still genuine and intense, it is found to have abated little of its claim to be obeyed." (2015, 11)

[37] Erich Hatala Matthes articulates a similar concern: "The problem is that the unstructured public does not generally offer a reliable mechanism for accountability, so even when cancel culture gets it right, it doesn't help to change or build institutions that offer the kind of accountability that would help to prevent future abuses." (2021, 101)

> threats of violence. Once public shaming begins, no one has the power to end it. Apologies from wrongdoers, no matter how well designed, are surprisingly ineffective in these cases. (2020, 54)

It is clear enough that informal social punishments can strike fear into many people and generate significant chilling effects. At this juncture, it is appropriate to consider how censorship ought to be conceptualized within a Millian framework. It goes without saying that censorship involves the suppression of speakers and the suppression of expressive content. However, reasonable people might disagree about whether the kinds of insidious social pressure that Mill warns against amount to a form of censorship, since this language often conjures imagery of state officials using force, and the threat of force, to shut down prohibited forms of expression. Instead of thinking of this kind of social pressure as something that is conceptually distinct from censorship, we may be better served by establishing a clear conceptual distinction between political censorship and social censorship.[38] Social censorship occurs when social punishments are successfully deployed in order to chill expression.[39] The informal nature of social censorship is what makes it distinct from political censorship, which is usually the focus of free speech theorists.

[38] Philosopher J.P. Messina also uses the language of social censorship and invokes Mill in his book *Private Censorship*: "Like Mill, when I speak of social censorship, I mean overt attempts to suppress speech by means of sanctions like naming, shaming, shunning, blaming, gloating dissociation, and so on." (2023, 53)

[39] Since this discussion deploys the language of intimidation and social punishment in addition to that of social censorship, it is worth clarifying the distinctions between these concepts. Within my framework, "social punishment" is a category that encompasses any form of action by private actors that seeks to pressure an individual or group into altering its conduct. Insults, ostracism, boycotts, and even threats of physical violence can all be understood as forms of social punishment, so long as they are carried out with the goal of exerting influence on the conduct of others. Accordingly, attacking people at random, with no discernable overarching objective, cannot properly be understood as a form of social punishment. "Social censorship" involves the successful deployment of social punishments in order to chill expression. While social punishments may achieve their objectives or fail to, social censorship by definition involves a successful attempt to silence speakers. When an individual or group abstains from an expressive act in order to avoid social punishment, social censorship is present. "Intimidation" is a process wherein individuals and groups are deliberately made to feel fearful by others. While intimidation can be a means of bringing about social censorship, it is not identical with social censorship. There are other forms of social censorship that do not necessarily involve fear. One example is the inundation of communities with spam messages in order to distract participants and derail discussion. While intimidation may be understood as the most common and efficacious means of achieving social censorship, the category of social censorship should be conceptualized as broad enough to include efforts to silence expression that deploy tactics besides intimidation.

While political censorship involves agents of the state constraining expression through codified prohibitions, restrictions, and punitive activities, the former unfolds in a manner that is more fluid.

Although intimidating people into silence using tactics such as ad hominem attacks, accusations of guilt by association, and strawmen are relevant examples of social censorship, social censorship should be understood as a broader category that encompasses a range of behaviours that extend well beyond these tactics. A clear conceptual distinction between political censorship and social censorship is helpful because it conveys two important points: the first is that free expression can be undermined by a broad array of actors that transcend the state. The second is that social dynamics are not equivalent to formal policies implemented by the state, and should not be conflated with them. These points are helpful for clarifying and substantiating the claim that despite the fact that social pressure does not (necessarily) involve state actors or the use of force, it can still amount to a serious threat to free expression in a Millian framework. While political censorship and social censorship are two distinct phenomena, it is nonetheless the case that both can be devastating to the atmosphere of free expression that Millian liberals wish to cultivate.

ii.ii: Free Expression and Supporting Institutions

The above discussion has argued that within a Millian framework, it is clearly the case that there are threats to free expression that go far beyond the scope of state power. We have noted that communities can injure free expression in a variety of ways without ever calling upon agents of the state to subject citizens to sanctions such as fines, incarceration, or corporal punishment. At this point, it is necessary to point out that due to the fact that institutions beyond the state can play a pivotal role in nurturing or thwarting an atmosphere of free expression, societies that wish to

nurture such an atmosphere may have positive responsibilities that go beyond merely permitting unpopular or controversial speakers to express themselves. This is because an atmosphere of free expression does not merely involve various individuals and groups having formal and informal permission to express themselves in an unfettered manner. It also involves individuals and groups with very different attitudes and worldviews making proactive efforts to engage with those who think and live differently from themselves. A society wherein people with different worldviews live free from censorship, but are heavily segregated and occupy discrete social siloes, will not give rise to the atmosphere of free expression for which Mill advocates in *On Liberty*.[40] In order for such an atmosphere to be present, institutions must be designed in a manner that is conducive to the good-faith exchange of ideas between societal actors that are at significant variance with one another.[41] A proper atmosphere of free expression requires supporting institutions that have the power to facilitate meaningful dialogue between actors who think differently.

While there are many ways in which institutions might be designed or reformed, there are two specific desiderata that deserve explication when we confront the question of how institutions can assist in supporting an atmosphere of free expression. The first is for institutions to be designed in a manner that enables members of the public to attain a certain baseline of competence with respect to their skills related to communication and debate. While some people are naturally gifted

[40] Mill argues that it is important for humans to be accustomed to diversity via coexistence and interaction with others who are different from themselves. He writes that in order for individuality to prosper, it is necessary for people " … to see that it is good there should be differences … even though, as it may appear to them, some should be for the worse." (2015, 72)

[41] Mill states that rival groups benefit from each other's contributions despite their antagonisms: "Individuals, classes, nations, have been extremely unlike one another: they have struck out a great variety of paths, each leading to something valuable; and although at every period those who travelled in different paths have been intolerant of one another, and each would have thought it an excellent thing if all the rest could have been compelled to travel his road, their attempts to thwart each other's development have rarely had any permanent success, and each has in time endured to receive the good which the others have offered." (2015, 71)

when it comes to articulating their beliefs and contemplating the beliefs of others, for most people, this is a skill that requires a considerable amount of training and practice. Setting aside questions of whether educational institutions ought to be administered publicly or privately (Mill calls for both types of institution to be permitted to coexist[42]), it is necessary to point out that an adequate atmosphere of free expression will involve proactive, organized efforts being made to equip members of the public with concepts and language that will enable them to enter into meaningful dialogue with others. A society that offers educational resources to all of its members, yet fails to use these resources in a manner that teaches people how to engage in productive intellectual inquiry will fail to cultivate a Millian atmosphere of free expression. Institutions must be designed with the goal of edifying students and preparing them to become active participants in a world wherein many people think differently than they do, and where available information will perennially be in flux. The goal is for people to have the competence, and not just the freedom, required to entertain a vast array of competing ideas, and use reason in order to identify which of these ideas are the soundest. This is a process that involves constantly being open to criticism as well as new information, and it is a far cry from situations wherein students are simply handed a set of conclusions that they are expected to memorize and endorse for the sake of advancement within a particular education system.

The second desideratum that must be fulfilled in order for a Millan atmosphere of free expression to thrive is the establishment of institutions wherein intellectual diversity is welcomed, and people cannot easily be punished merely for challenging orthodoxies and entertaining ideas

[42] Regarding the state's role in education, Mill states: "An education established and controlled by the State should only exist, if it exist at all, as one among many competing experiments, carried on for the purpose of example and stimulus, to keep the others up to a certain standard of excellence." (2015, 103)

that are unpopular. Freedom to share one's views with others becomes moot in situations wherein institutions are intolerant of dissent and are quick to expel participants who challenge the reigning orthodoxy of the day. It is critical to note that this undermines an atmosphere of free expression even when those ejected from institutions are permitted to seek sympathetic audiences elsewhere and are given opportunities to commune with the like-minded. Even if alternative institutions exist that can welcome people who have been targets of expulsion, it is nonetheless true that an atmosphere of free expression is undermined when individuals and groups with different views are forced to inhabit rival institutions, and opportunities for interaction and debate between those with different worldviews are thereby curtailed. Simply put: an atmosphere of free expression demands not only diversity between institutions, but also diversity within institutions. The former kind of diversity lends itself to complacency and monotony by encouraging people to focus on pleasing members of their in-group while ignoring members of their out-group. Meanwhile, the latter kind of diversity lends itself to intellectual refinement and progress[43] as people are incentivized to formulate ideas and arguments that can appeal to others who do not already share their intellectual priors, meaning that they must grapple with challenging questions and criticisms in order to strengthen the case on offer.

Some caveats must be acknowledged here. It would be wrongheaded to suggest that every institution must function as a welcoming venue for every type of person and every type of worldview. This is an extreme prescription that would throw many kinds of institutions into

[43] According to Mill, people frequently arrive at a clearer understanding of the truth by synthesizing rival views on an issue: "We have … considered only two possibilities: that the received opinion may be false, and some other opinion, consequently, true; or that, the received opinion being true, a conflict with the opposite error is essential to a clear apprehension and deep feeling of its truth. But there is a commoner case than either of these; when the conflicting doctrines, instead of being one true and the other false, share the truth between them; and the nonconforming opinion is needed to supply the remainder of the truth, of which the received doctrine embodies only a part." (2015, 45)

disarray. In many cases, institutions must put in place rules, requirements, and admissions criteria in order to maintain a reasonable level of cohesion and reliability. The point is that those who wish to question and revise the status quo inside institutions should not be punished or thrown out, formally or informally, merely for voicing their ideas and criticisms. Good-faith efforts to expose weaknesses in the patterns of thought and behaviour that pervade institutions can be of enormous benefit, and ought not be blocked or sanctioned. Institutions can provide space for debate and dissent without derailing their essential functions. Importantly, while these prescriptions can be applicable to many types of institutions, they are most relevant with respect to institutions that are explicitly committed to intellectual inquiry and the advancement of knowledge. Institutions of this kind ought to provide ample room for the expression and discussion of controversial and unpopular ideas, as this kind of probing is precisely what enables people to arrive at a clearer understanding of the truth. If people inside of institutions raise concerns or criticisms that are ill-considered or out of line with reality, then they should be criticized and rebutted rather than expelled. The notion that some ideas must be blocked from discussion because they are too wrong to merit consideration is self-defeating, as ideas of this nature should be easily refutable by those who are better equipped in terms of facts and reason.

Beyond merely tolerating criticism and debate, institutions can and should provide spaces wherein this type of communication is actively encouraged. The ability to grapple with diverse and challenging ideas is one that is best developed when people make its use part of their routine. It is desirable for people to become accustomed to the process of addressing disagreements directly and explicitly, without interpersonal animus or excessive displays of emotion. For example, institutions can help normalize this kind of communication by establishing specific times and

places wherein personnel are expected to engage in it. Even when no strong disagreements are present between participants in an institution, such organized efforts send a powerful signal about the value of intellectual diversity, thereby encouraging original thought and helping to secure an atmosphere of free expression. It is one thing to permit intellectual diversity and it is another to actively invite it, and the latter is necessary in order for the intellectual climate of institutions and societies to be optimized.

It is important to acknowledge that there is of course a significant gap between outlining the kinds of institutions that are needed in order to establish an atmosphere of free expression, and actually effecting social change so that such institutions can come into being. Societies that perpetuate subordination, marginalization, and exclusion of particular populations via their institutional arrangements will need to go through a process of evolution before they can erect institutions that meet the criteria outlined above. This is because in order for meaningful communication between diverse populations to take place, a certain baseline of mutual social recognition must first be attained. Philosopher Rae Langton argues that in cases wherein severe power imbalances exist between individuals and groups, an ability to speak with ostensible freedom may not entail an ability to be understood and treated in a manner that is conducive to uptake:

> The ability to perform speech acts of certain kinds can be a mark of political power. To put the point crudely: powerful people can generally do more, say more, and have their speech count for more than can the powerless. If you are powerful, there are more things you can do with your words … If you are powerful, you sometimes have the ability to silence the speech of the powerless. One way might be to stop the powerless from speaking at all. Gag them, threaten them, condemn them to solitary confinement. But there is another, less dramatic but equally effective, way. Let them speak. Let them say whatever they like to whomever they like, but stop that speech from counting as an *action*. More precisely, stop it from counting as the action it was intended to be … it is a kind of silencing about which Austin had something to say, without commenting on its political significance. Some

> speech acts are *unspeakable* for women in some contexts: although the appropriate words can be uttered, those utterances fail to count as the actions they were intended to be. (2009, 30-31)

Langton goes on to state: "Free speech is a good thing because it enables people to act, enables people to do things with words: argue, protest, question, answer. Speech that silences is bad, not just because it restricts the ideas available on the shelves, but because it constrains people's actions". (2009, 61) This view underscores the idea that societies that are afflicted by intense inequality and division will struggle to design supporting institutions that bolster free expression. Societies wherein the perspectives of certain demographics are dismissed as a matter of course, and pushed to the margins, will of course fall short of developing mainstream venues wherein discourse can proceed with fairness and openness.[44] Messina emphasizes the importance of interlocutors viewing one another as peers in order for the values associated with free expression to flourish:

> …my ability to be heard by others depends on more than [its] being the case that they're around when I choose to speak. It also requires that I have a certain kind of standing in the community. If I speak on some matter of concern to an audience that does not hold me to be an epistemic peer, then in some real sense my words fall on deaf ears. This brings home how important the presumption of a good reputation is for a public sphere that is likely to realize the value of free speech. (2023, 35)

While making a case for institutions that are supportive of intellectual diversity is relatively simple and straightforward, cultivating social relations wherein the establishment of such institutions is feasible is much more complex. Accordingly, it is necessary to point out that for some societies, the process of designing institutions that meet the aforementioned criteria is one that must be conceptualized as a long-term project that will involve broader cultural change. The question of

[44] Messina states: "In broad outlines we want an environment that encourages wide participation across social groups. A social context so replete with derogatory speech against women or minorities that they check out of the conversation or are not heard when they attempt to participate is one that will likewise fail to uncover, properly diagnose, and adequately respond to social problems." (2023, 32)

how to achieve such cultural change is too rich to treat in detail here, but it is nonetheless appropriate to flag its relevance and importance.

ii.iii: Free Expression and Social Goods

Thus far, this discussion has sought to outline a normative vision of free expression with two essential components. One of these components is largely negative, and consists in the absence of social and political pressures that are intended to silence disfavoured expression. The other is largely positive, and consists in the erection of institutions that support the productive exchange of ideas throughout society between individuals and groups with different worldviews. At this juncture, it is appropriate to say more about why free expression is desirable in the first place. Rather than further discussing what is demanded by a Millian system of free expression, we will now explore some of the features of free expression that make it a net benefit to society. The aim is to offer arguments about the value of free expression that might prove persuasive to people who are skeptical about the ability of free expression make societies better off. What can Millian liberals say to those who fear that free expression corrodes society by enabling the perpetuation of various forms of expression that have no redeeming qualities?

Throughout the rest of this chapter, it will be argued that free expression helps to promote three important social goods.[45] These social goods are critical intellectual faculties, authenticity in discourse, and equity in accountability. Much more will be said about these in the discussion that follows. For now, let us take a moment to consider what exactly a social good involves. For the

[45] It is noteworthy that Sumner also uses the language of "social goods". In *The Hateful and the Obscene*, he states: "It must … be shown … that criminalizing hate speech will succeed in reducing its circulation, with corresponding gains in self-esteem and other important social goods for the members of target minorities." (2004, 64) In Sumner's article in the book *Extreme Speech and Democracy*, the same line appears. (Hare and Weinstein, ed., 2009, 206)

purposes of this discussion, a social good can be understood as any feature of community life that reliably advances utility, or human happiness, as well as the resilience of society over the long term.[46] Both of these criteria must be met in order for a feature of community life to qualify as a social good. A feature of community life that generates significant gains with respect to people's happiness, but simultaneously lays the groundwork for societal dysfunction and decay, cannot qualify as a social good. While we can imagine many forms of behaviour that produce pleasure in the short term, it would be inappropriate to describe these as social goods when they undermine the long-term tenability of society. Alternatively, a feature of community life that binds a society's people closer together and promotes solidarity, but also immiserates them, will likewise fail to qualify as a social good for our purposes. For the purposes of this discussion, we are interested in features of community life that augment human happiness while also bolstering the ability of human beings to exist peacefully alongside one another in large networks for years, decades, centuries, and so on.

The arguments presented in the sections that follow are clearly inspired by Mill's work in political philosophy, but the aim of this discussion is not to engage in interpretive argumentation. Rather, the aim is to locate conceptual tools that can be helpful for understanding why an atmosphere of free expression is important and worthy of protection, and also the many ways in which it can be jeopardized. Since later chapters will explore ways in which certain aspects of contemporary public discourse are injurious to free expression, the discussion of these three social

[46] It is clear from Mill's *Utilitarianism* that that his goal is to maximize utility over the long term, and that this project involves getting people to internalize the idea that their own good is bound with the good of others: "… utility would enjoin, first, that laws and social arrangements should place the happiness, or (as speaking practically it may be called) the interest, of every individual, as nearly as possible in harmony with the interest of the whole; and secondly, that education and opinion, which have so vast a power over human character, should so use that power as to establish in the mind of every individual an indissoluble association between his own happiness and the good of the whole…" (2015, 131)

goods can aid us in understanding precisely what is at stake in this context. It is worth noting that in highlighting the social goods of critical intellectual faculties, authenticity in discourse, and equity in accountability, I am not suggesting that these are the only social goods that are associated with free expression. There is plenty of room for further conversation and debate about the social goods that free expression helps to promote, and I do not pretend to have provided an exhaustive list in this chapter. These specific social goods are brought to the fore in this discussion because they are closely associated with Mill's primary texts, and also because they are highly relevant to debates about modern public discourse and its dysfunctions, which will occupy our focus in later chapters.

ii.iv: Critical Intellectual Faculties

Critical intellectual faculties are the cognitive tools that humans use to assess the strengths and weaknesses of ideas. Whenever an individual makes an effort to deploy reason in order to examine an idea and reach a conclusion about whether it is capable of surviving scrutiny, they are deploying their critical intellectual faculties. Just as one might apply pressure to a piece of physical material in order to evaluate its strength, critical intellectual faculties enable people to apply pressure to claims and arguments so that their strength can be evaluated. This of course does not mean that every conclusion that is reached via a process of critical thinking is necessarily correct. People can deploy their critical intellectual faculties, and do so in earnest pursuit of the truth, and still end up endorsing conclusions that are flawed. However, it is nonetheless true that critical intellectual faculties have the ability to function as safeguards that prevent individuals and groups from accepting ideas that are unsound. These faculties are a powerful tool for filtering out falsehoods and aiding people in the project of determining which ideas among a set of alternatives

are the most sound. This is desirable, especially insofar as acceptance of falsehoods can impel people to behave in ways that are detrimental to themselves and others.

Rather than conceptualizing critical intellectual faculties as goods that are either present or absent among particular individuals, we can do better by thinking of these faculties as existing on a continuum. Some people possess exceptionally weak critical intellectual faculties, and others possess exceptionally strong critical intellectual faculties. What matters most for our purposes is that these faculties can be stunted or nurtured depending on the environment in which individuals and groups find themselves. Individuals and groups that are routinely given opportunities to assess competing ideas will generally do a better job of developing their critical intellectual faculties because they will be more practiced in the process of identifying the internal consistency of ideas, as well as their compatibility with available evidence. Like countless other domains, the domain of critical thinking is one wherein people become more skillful through experience. When people are actively encouraged to consider ideas and formulate critiques of them, we can expect their critical intellectual faculties to thrive, especially when they are given ample opportunity to be exposed to critiques of their own position. The dual process of critiquing and being critiqued can aid people in distinguishing between sound and unsound ideas so that their worldview can become more accurate over time.

Alternatively, people's critical intellectual faculties can be stunted when they find themselves in a setting wherein the process of exploring and critiquing ideas is disincentivized. When orthodoxies are established and people are instructed not to question prevailing beliefs, this means that opportunity for the deployment of critical thought is constrained. Social censure can be

a powerful means of chilling discussion and ensuring that orthodox beliefs are protected from scrutiny. In such scenarios, even if people continue to privately question orthodox beliefs and apply their critical thinking skills to them, they will be deprived of the opportunity to engage in dialogue with others that could help deepen their understanding of the relevant issue. When social sanctions for critical thought and discussion are present, we can expect sizeable portions of the relevant population to engage in self-censorship in order to avoid triggering these sanctions. This means that discussion is stifled, as well as the critical intellectual faculties of those who might profit intellectually from the opportunity to critique and be critiqued in the manner described above.

Of course, Mill has much to say about how freedom from conformist pressures is favourable to the intellectual development of human beings. He views free expression as a mechanism that enables humans to develop their intellectual faculties through engagement with individuals and groups with alternative views that can challenge their own.[47] Echoing the biblical saying that "as iron sharpens iron, so one person sharpens another", Mill's discussion of the value of free expression highlights the importance of people confronting the strongest arguments that can be marshalled against their own positions in the interest of revising their worldview so that it can become more clear and refined over time. Free expression raises opportunities for all sorts of

[47] The following passage illustrates Mill's idea that considering views on all sides of an issue is crucial for the development of critical intellectual faculties: "Why is it, then, that there is on the whole a preponderance among mankind of rational opinions and rational conduct? … it is owing to a quality of the human mind, the source of everything respectable in man either as an intellectual or as a moral being, namely, that his errors are corrigible. He is capable of rectifying his mistakes, by discussion and experience. Not by experience alone. There must be discussion, to show how experience is to be interpreted. Wrong opinions and practices gradually yield to fact and argument; but facts and arguments, to produce any effect on the mind, must be brought before it. … In the case of any person whose judgment is really deserving of confidence, how has it become so? Because he has kept his mind open to criticism of his opinions and conduct … Because he has felt, that the only way in which a human being can make some approach to knowing the whole of a subject, is by hearing what can be said about it by persons of every variety of opinion, and studying all modes in which it can be looked at by every character of mind…" (2015, 22)

individuals and groups to have their ideas challenged, meaning that flaws in reasoning can be exposed and erroneous beliefs can eventually give way to beliefs that are more sound. Mill holds the view that being wrong, and having one's wrongness brought to the fore by interlocutors, can be a valuable component of the search for truth,[48] and that the difficult process of trial and error that accompanies freewheeling discussion and debate is far more beneficial to the human intellect than rote education. We might distill Mill's position by saying that while rote education can help make people more knowledgeable, it fails to make people wiser, as wisdom is a virtue that is best cultivated in an atmosphere of freedom wherein diverse individuals and groups can communicate openly without fear of punishment for exploring or defending ideas that are taboo.[49]

Mill views a restrictive intellectual atmosphere as a type of fetter that prevents human beings from reaching their full potential, arguing that when society punishes actors for espousing views that have been deemed unacceptable, ordinary people's "whole mental development is cramped, and their reason cowed, by the fear of heresy." [50] To live in a society wherein mental freedom is denied is effectively to live in an arrested state wherein one's intellectual faculties are not fully developed, which means that they cannot be fully capitalized on by society. The sensibility that animates this argument is that the human mind, while not a muscle in a literal sense,

[48] The following highlights the importance of giving people space to think independently, even when this independent thought may involve errors: "… the peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error." (2015, 19)

[49] While critical intellectual faculties and wisdom are not identical, as the former may be deployed in order to advance ideas and agendas that are decidedly unwise, it is nonetheless true that critical intellectual faculties function as a conduit towards wisdom as they enable people to rigorously assess the grounds for specific conclusions and make rational judgments as to whether a given conclusion is justified. Alternatively, rote education constrains the process of reason-giving and rewards individuals for conforming to intellectual orthodoxies rather than for questioning them.

[50] The following conveys Mill's view that constraints on inquiry are damaging to people's intellectual development: "... it is not the minds of heretics that are deteriorated most by the ban placed on all inquiry which does not end in the orthodox conclusions. The greatest harm done is to those who are not heretics, and whose whole mental development is cramped, and their reason cowed, by the fear of heresy." (2015, 34)

is analogous to a muscle in that it needs to be used and challenged in order to develop appropriately. Being exposed to many competing ideas on a subject and having to reach one's own conclusions is often a challenging experience, but it is a challenge worth undergoing, as it is conducive to individuals' and societies' intellectual progress.

While Mill is a staunch defender of intellectual diversity and freewheeling debate, it is crucial to emphasize that the Millian worldview does not imply that all ideas are equally valid or worthy of respect. Rather, it implies that instead of punishing speakers for endorsing wrongheaded beliefs, we ought to channel our energies towards challenging the beliefs themselves via argumentation. A culture of open inquiry and rigorous debate can be a powerful filtering mechanism. Rather than establishing political (formal) or social (informal) constraints with respect to expression that are designed to prevent people from thinking and speaking in a manner that is wrong, we are better off in the aggregate by permitting people to commit errors in their thought and speech, with the expectation that other individuals and groups will check them and balance out their erroneous perspectives with alternative perspectives. Moreover, institutions can facilitate and accelerate this process by creating spaces wherein this kind of intellectual competition is encouraged. Even if none of the parties engaged in dialogue are entirely correct, the perpetual process of checking can limit the spread of pernicious ideas and ensure that individuals and groups are not incentivized to reach extreme positions as a result of interacting exclusively with actors that adhere to a specific worldview. It of course must be conceded that a Millian approach to free expression does create space for speech that is unproductive, such as insults, gossip, and bad-faith arguments.[51] However, it also creates space for critical assessment of these forms of expression. If

[51] It must be noted that even though Mill is the most prominent advocate for free expression in the philosophical canon, it would be an error to view him as an absolutist with respect to this topic. Mill's harm principle lays the groundwork for identifying cases wherein it is legitimate for society to use force to put a stop to expressive acts. While Mill

Mill is right that free discussion strengthens people's critical intellectual faculties, then we can expect them to gradually develop an aversion to forms of expression that are petty and unproductive and to seek out arguments that genuinely strive to apprehend important truths, so long as meaningful discussion is permitted to carry on without interruption.

There is another case for the importance of critical intellectual faculties that ought to be appreciated here. In *Utilitarianism*, Mill makes clear that while he believes that the overarching ethical objective of our conduct should be to maximize pleasure and minimize pain, it does not follow from this that all forms of pleasure are equally valuable.[52] He makes the case that pleasures that involve higher faculties are inherently more meaningful than ones that are merely sensual in nature. Accordingly, when individuals and societies undergo a process whereby their powers of critical thinking are strengthened, they open up new opportunities for pleasure that had previously been unavailable to them. The development of critical intellectual faculties is thus important from a Millian perspective on multiple levels. In addition to helping to steer society, and the individuals that comprise society, away from erroneous ideas and towards sound ones, critical intellectual faculties enable human beings to experience higher forms of pleasure. It is thus appropriate to

endorses the view that all individuals ought to have broad freedom to express themselves, even when the ideas that they espouse are widely or intensely disliked, he also notes that in some scenarios, expressive acts can generate significant harms, and can therefore be legitimately sanctioned. He explains that "…even opinions lose their immunity when the circumstances in which they are expressed are such as to constitute their expression a positive instigation to some mischievous act … Acts, of whatever kind, which, without justifiable cause, do harm to others, may be, and in the more important cases absolutely require to be, controlled by the unfavourable sentiments, and, when needful, by the active interference of mankind. The liberty of the individual must be thus far limited; he must not make himself a nuisance to other people." (55) While there is no consensus regarding precisely which expressive acts are legitimate targets of punishment, there is generally broad agreement among classical liberals influenced by Mill that certain categories of expression, such as fraud, threats, and defamation, are inherently illegitimate and should therefore be legally prohibited.

[52] Mill states: "It is quite compatible with the principle of utility to recognise the fact, that some kinds of pleasure are more desirable and more valuable than others. It would be absurd that while, in estimating all other things, quality is considered as well as quantity, the estimation of pleasures should be supposed to depend on quantity alone." (2015, 122) He goes on to explain that "… it is an unquestionable fact that those who are equally acquainted with, and equally capable of appreciating and enjoying, both, do give a most marked preference to the manner of existence which employs their higher faculties." (2015, 123)

conclude that the gains that accompany the cultivation of these faculties are multifold. Since there are certain types of pleasure that are inaccessible when critical intellectual faculties are stifled, this gives us an additional basis for recognizing their importance from a Millan perspective.

We have seen that within a Millian framework, critical intellectual faculties are an important social good that can be undermined in various ways. While this is true of both political and social censorship, the latter form of censorship is most relevant for our purposes, and it is also the one that Mill views as most insidious. Despite the fact that it does not invoke state power, social censorship blocks people from making worthwhile contributions to society by questioning entrenched thinking and engaging in thoughtful dialogue with others. If we are interested in constructing a society that can be resilient in the face of various threats and challenges, be they internal or external, then we have strong grounds to minimize social censorship and the conformist pressures that it entails so that each member of society can have a meaningful opportunity to reach their full developmental potential. This dynamic enables society to maximize the amount of intellectual capital available to it. If Mill is right that an atmosphere of free expression helps to promote the social good of critical intellectual faculties, then we have strong grounds to cultivate such an atmosphere, which may involve constructing institutions that explicitly call for the deployment of critical intellectual faculties with respect to a broad array of ideas and propositions.

ii.v: Authenticity in Discourse

If we accept, in Millian fashion, the idea that an atmosphere of free expression can be productive at the societal level because of its ability to expose falsehoods and exchange them for truths, then it is appropriate to say something about what exactly constitutes productive public discourse, and how productive public discourse might be corrupted. Unless we want to accept the

idea that all public utterances are invariably a social good, which seems wrong on its face,[53] then it is incumbent upon us to develop an account of what type of public discourse we wish to cultivate. An idea I wish to advocate for, and one that is inspired by Mill,[54] is that a key component of productive public discourse is authenticity: people saying in public what they genuinely think as private individuals. Authenticity is absent from public discourse when people say in public what others want to hear because they have succumbed to conformist pressures. Public discourse becomes frivolous if it is animated by a desire to impress others and garner social currency rather than a desire to identify and promote ideas that are sound; if public discourse is driven merely by a desire for social approval, then it is plain to see how conformity can run rampant and the actual soundness of ideas can become an afterthought.[55] While it is true that in some cases engaging authentically with others may involve certain types of social rewards, in many cases it will involve significant discomfort, and in such cases, authenticity ought to take precedence over the accumulation of social rewards.

In order to appreciate the value of authenticity, it is worthwhile to consider the related concept of timidity. By posing a rhetorical question, Mill notes that timidity is a common feature of human beings that can prevent them from making worthwhile contributions to their communities and the world at large: "Who can compute what the world loses in the multitude of

[53] For example: it is difficult to see how people directing petulant insults at one another amounts to a social good, even if we accept that it amounts to a form of public discourse and that the freedom to engage in such discourse is important to preserve.

[54] Mill links "non-conformity" and "eccentricity" with the virtues of "mental vigour" and "moral courage". In addition, he laments the fact the only a small minority of people are willing to be eccentric: "In this age, the mere example of nonconformity, the mere refusal to bend the knee to custom, is itself a service. Precisely because the tyranny of opinion is such as to make eccentricity a reproach, it is desirable, in order to break through that tyranny, that people should be eccentric. Eccentricity has always abounded when and where strength of character has abounded; and the amount of eccentricity in a society has generally been proportional to the amount of genius, mental vigour, and moral courage which it contained. That so few now dare to be eccentric, marks the chief danger of the time." (2015, 66)

[55] Hrishikesh Joshi articulates this point: "The aggressive conformist is primarily motivated by winning the zero-sum game of social status. Like the contrarian or the troll, he is not motivated by promoting the health of the epistemic commons or trying to reach the truth on some issue." (2021, 82)

promising intellects combined with timid characters, who dare not follow out any bold, vigorous, independent train of thought, lest it should land them in something which would admit of being considered irreligious or immoral?" (2015, 34) This question implies that even in a society wherein many people possess significant intellectual gifts, discourse can be impoverished when these individuals are prevented from showcasing to others the thoughts and arguments that they have to offer out of fear of being targeted with censure and ostracism. It is an unfortunate fact, in Mill's view, that timidity is a pervasive quality that robs society of the benefits associated with heterodox discussion and debate.

Let us accept as a premise that incentives can shape public discourse, just as they can shape countless other aspects of human behaviour. A concerning feature of censorship is its impact on the incentives that are at play in public discourse with respect to authenticity. Social censorship puts pressure on individuals and groups to espouse ideas not because they sincerely believe that the ideas are correct, but because they wish to avoid social punishment and remain in the good graces of others. The more that members of a society are pressured to espouse a certain set of beliefs, the less confidence we can have that their expressed commitment to said beliefs is genuine. Why should people communicate in an authentic manner when the downsides far outweigh the benefits? Why should people publicly challenge entrenched orthodoxies when this will jeopardize their standing in their community? While certain individuals who are exceptionally courageous may publicize their authentic views when doing so can be very costly to them, in most cases, we can expect people to conceal these views in order to protect their reputation and avoid the scorn of their fellow citizens.

Mill's arguments in *On Liberty* provide insight about the danger of inauthenticity in public discourse. A key idea that runs throughout Mill's writings is that when we use "social intolerance" to banish ideas that are deemed unacceptable, this generally fails to bring about meaningful change with respect to people's views. [56] Instead, social intolerance impels people to "disguise" their views so that they can be protected from the negative judgment of the public. Mill tells us that even when the forces of social censorship within a society are strong, prohibited discussions will continue to take place, and prohibited beliefs will continue to be held, away from the view of the public. There is of course an important difference between the elimination of beliefs and the concealment of beliefs. When societies dole out social punishment for the expression of prohibited views, this increases the likelihood that people will be deceived into believing that their society has been cleansed of certain views, when in reality these continue to be perpetuated in "narrow circles" wherein they do not face robust scrutiny because alternative viewpoints are absent.

Let us accept, as most people do, that some ideas are objectively better than others with respect to their soundness. Let us also accept that it is legitimate to want to disabuse society of pernicious ideas and supplant them with superior ones. The important point for our purposes is that it does not follow from these premises that social censorship is the appropriate path forward. Indeed, Mill's arguments suggest that the opposite is true: social censorship makes the project of filtering out bad ideas more difficult by incentivizing people to hide these ideas from public venues where they can be rigorously analyzed and challenged. In a Millian framework, if we want to disabuse society of pernicious ideas, what we ought to do is cultivate an atmosphere of free expression so that these ideas can be confronted, challenged, and ultimately supplanted through a

[56] Mill states: "Our merely social intolerance kills no one, roots out no opinions, but induces men to disguise them, or to abstain from any active effort for their diffusion." (2015, 33)

process of unrestricted discussion and debate. If Mill is correct, as I think he is, in arguing that in many cases social censorship merely drives ideas underground instead of supplanting them, then the notion that chilling expression is warranted for the sake of eliminating bad ideas from society necessarily falls short, as it proceeds from an erroneous premise.

The above comments point to the conclusion that authenticity in public discourse is a valuable social good. It is valuable because it enables members of a society to communicate meaningfully with one another, and to identify areas of disagreement rather than conceal them. Instead of espousing the (clearly wrong) view that concealing ideas from the public is disadvantageous because all ideas necessarily have merit, I advance the view that concealment of ideas is problematic because it prevents members of society from knowing that pernicious beliefs are circulating around them, and accordingly undermines their ability to work towards combating them. By definition, social censorship pressures people into silence, sending a message that certain views and topics are so beyond the pale that it is illegitimate for them to be examined. If pernicious ideas are not examined, then they cannot be rebutted. Inauthenticity distorts our understanding of our own community and society, and prevents us from filtering out bad ideas through rigorous discussion and debate. This is why we have strong grounds for constraining social censorship, and creating space for the social good of authenticity in discourse to flourish.

Perhaps even more importantly, authenticity in discourse ought to be conceptualized as a social good because people need it in order to forge meaningful relationships and communities. If people are not free to communicate in a manner that is authentic, then their social connections will lack integrity and robustness: rather than conversing with others and arriving at agreements and consensuses that are genuine, people can instead be expected to follow the proverbial herd and pay

lip service to ideas that they do not genuinely believe in or even understand. A society that lacks authenticity in discourse will struggle to cultivate trust, and its people and institutions will be viewed with unease and suspicion. In severe cases, this could lay the groundwork for social and political conflict that eventually transcends words and involves the use of force. If we are at all interested in cultivating a society wherein social ties are robust and resilient, then we have good reason to care about authenticity in discourse and its ability to forge social connections and communities that can thrive over the long term. In situations wherein social connections and communities are held in place by the pressures of conformity rather than genuine and substantive deliberation, we can expect these relationships to deteriorate or even perish when this pressure is lifted or when competing pressures point in a different direction. Inauthentic public discourse leads to fickle social relations that can easily wither away if and when people sense that they may secure more social rewards by aligning themselves with a different assortment of people and ideas. If we wish to avoid the cheapening of all social life, authenticity in discourse is a good that ought to be preserved.[57]

While this discussion of authenticity in discourse has emphasized its social value, it would be an error to overlook its profound value for individuals. Just as critical intellectual faculties can enrich the lives of individuals in addition to enriching the intellectual life of a society at the macro level, authenticity in discourse can go a long way in making life more fulfilling and enjoyable for the population at large. For the vast majority of people, there is something deeply uncomfortable

[57] Economist Glenn C. Loury describes how self-censorship can damage social ties: "The risk of self-censorship ... is not so much that our public intellectuals and political leaders will be repressed, but that private citizens will be. If we cannot know what our friends, family, neighbors, and community truly think because they fear reprisal, we cannot know ourselves. Social life abhors a discursive vacuum ... silence, enforced by fear, will be filled with suspicion, betrayal, and the shattering of social bonds." (2025, 78)

and stressful about the experience of having to conceal one's genuine thoughts. This dynamic can take a toll on people's mental and emotional wellbeing for a variety of reasons. In addition to calling into question the integrity of their relationships with others, such a life of inauthenticity can inflict damage on people's wellbeing by generating feelings of confusion and alienation as they struggle to reconcile the beliefs that they profess in public with the beliefs that they embrace in private and intimate settings. As a result of this incongruence, individuals may succumb to the view that the version of themselves that exists away from the public is shameful in some way, and accordingly experience unjustified feelings of guilt and anxiety. Ironically, in some cases, it may turn out to be the case that there are many other members of one's community who have similar thoughts and beliefs, but this common ground cannot be identified because social pressure requires each party to remain silent out of fear of attracting punishment. While inauthenticity in discourse is detrimental to society, it is also detrimental to the inner life of countless individuals who could benefit from opportunities to express their genuine views among others who share them, or among others who are at least willing to give them a fair hearing. If we are interested in maintaining a populace that is happy and at ease rather than one that is perpetually weighed down by misery and angst, then we ought to protect and promote the good of authenticity in discourse. This can be achieved by creating and sustaining a Millian atmosphere of free expression wherein people are permitted and encouraged to engage with interlocutors and ideas in an authentic manner.

ii.vi: Equity in Accountability

Thus far, this discussion has invoked the concept of social punishment in order to develop an account of how dynamics that take place beyond the purview of government can undermine an atmosphere of free expression and deprive societies of important social goods. At this point, it is

important to note that some forms of social punishment are far harsher than others, and therefore are more likely to produce intimidation among their targets. For example: while disdain and ridicule are properly understood as social punishments in their own right, they are generally less likely to instill fear in people than ostracism, or the severing of social ties. There are good reasons for this.[58] In cases wherein people are faced with severe ostracism after having fallen afoul of social norms, they may find themselves struggling to meet the basic requirements of life, such as feeding, sheltering, and clothing themselves. For the vast majority of people, generating income is a project that involves ample cooperation with others, which can include remaining in the good graces of employers and potential employers. Unless one is a citizen of a society with an exceptionally generous welfare apparatus, they must rely on their interpersonal relationships in order to achieve a decent standard of living and avoid a life of poverty. By definition, ostracism involves the dissolution of such interpersonal relationships. While it would be an exaggeration to say that social punishments invariably jeopardize people's physical wellbeing, it is fair to say that in addition to depriving people of many important sources of meaning and enjoyment in life, social punishments can even jeopardize their ability to fulfill their basic material needs. This dynamic merits serious consideration.

In addition to bringing attention to this matter in *On Liberty*, Mill notes that there is significant inequity built into this dynamic. Mill states the following:

> For a long time past, the chief mischief of the legal penalties is that they strengthen the social stigma. It is that stigma which is really effective, and so effective is it, that the profession of opinions which are under the ban of society is much less common in England than is, in many other countries, the avowal of those which incur risk of judicial

[58] Rauch notes that adaptive pressures have shaped how humans come to accept and reject beliefs: "You might think that perverse stubbornness would be maladaptive from an evolutionary point of view. The reason it is not goes back to Aristotle: humans are social animals. What matters most from an evolutionary perspective is not that a person forms beliefs which are true; it is that she forms beliefs which lead to social success." (2021, 30)

> punishment. In respect to all persons but those whose pecuniary circumstances make them independent of the good will of other people, opinion, on this subject, is as efficacious as law; men might as well be imprisoned, as excluded from the means of earning their bread. Those whose bread is already secured, and who desire no favours from men in power, or from bodies of men, or from the public, have nothing to fear from the open avowal of any opinions, but to be ill-thought of and ill-spoken of, and this it ought not to require a very heroic mould to enable them to bear. (2015, 32-33)

These comments draw attention to the fact that while it is true that any member of society may be a target of disapproval and hostility as a result of transgressing social norms, the impact of social punishment can vary greatly depending on the resources that its target possesses. When Mill states that people "might as well be imprisoned, as excluded from the means of earning their bread", this indicates that the loss of economic opportunities can be extremely damaging to a person's liberty. His subsequent statement that "[t]hose whose bread is already secured, and who desire no favours from men in power, or from bodies of men, or from the public, have nothing to fear from the open avowal of any opinions…" indicates that economic privilege can insulate certain individuals and groups from the chilling effects of social punishment, while others are much less fortunate. A person who leads a life of affluence has relatively little to fear when it comes to the issue of social intolerance, as their economic privilege guarantees that they can continue to live comfortably even if their reputation receives serious damage. While an affluent person may be cut off from worthwhile opportunities as a result of falling afoul of social norms, they have relatively little reason to worry in comparison to someone who earns a modest living and cannot easily cope with a sudden loss of income.[59]

What we have identified here is an important inequity that relates to the issue of social censorship, which includes, but is not limited to, behaviour that seeks to intimidate individuals and

[59] In his discussion of cancellation in the art world, Matthes states: "…we're left with the unsettling conclusion that the costs of cancel culture have primarily been borne by everyday people, including aspiring artists without fame or fortune to fall back on." (2021, 102)

groups into silence. If Mill is correct that social intolerance can have disproportionate chilling effects depending on the level of economic advantage that its targets possess, then this means that the phenomenon of social censorship is likely to be far more stifling towards the words and ideas of certain classes of people than others. If it happens to be the case that people with less economic privilege tend to espouse different views than those with greater economic privilege, then the latter set of ideas will be far more likely to be heard and considered than the former. While the powerful and well-connected can speak relatively freely, knowing that any blowback that they experience will not threaten their material security, people who occupy more vulnerable positions in society must exercise caution and make sure that they do not transgress a boundary that could seriously jeopardize their ability to meet their material needs.

It is of course important for people to be accountable for their words and conduct. Sometimes people speak and behave in a manner that is damaging to others, and it is vital that others have opportunities to challenge their conduct and demand improvements to it. All members of society, from the most privileged to the least, should face a reasonable level of accountability for their words and actions. However, if we construct a society wherein social punishments are deployed as our default method for achieving accountability, then we will likely fuel inequity, regardless of whether this is our intention. The reality is that the sting of social punishment is far more painful to some people than others, and this differential impact is likely to generate dynamics in public discourse that are far from equitable. Fortunately, we can preserve a commitment to accountability without jumping to the conclusion that social punishment is the best means of attaining it, as other strategies are available.

Perhaps instead of excluding those who espouse views that are determined to be pernicious, and driving them away from polite society, we ought to explore the possibility of making greater efforts to include them.[60] The reasoning here is that by making a concerted effort to expose such individuals to alternative perspectives and the people who hold them, we can help disabuse people of noxious views and help them develop social ties to individuals and groups that can act as positive intellectual influences. Even if people disagree about whether inclusion or exclusion is appropriate in a given case, the point is that alternatives to exclusion do exist, and we ought to be mindful of this if we are interested in preserving equity in accountability. There is really no reason why accountability for poor conduct should involve the severing of social ties in all cases. We must remain cognizant of the possibility that isolating people who offend others with their words and actions can make matters worse,[61] and that striving to forge connections with them can make matters better.[62] The forging of connections aligns with the Millian goal of cultivating dialogue between societal actors with different views, and can be beneficial with respect to the inner life of individuals, as was noted in the previous section.

There are several ways in which this project of inclusion might be realized in more concrete terms. Philosopher Elizabeth Anderson and political scientist Yascha Mounk have both advocated

[60] Redstone and Villasenor draw attention to the social benefits of being exposed to a broad array of divergent perspectives: "…contact on its own, even in the absence of the other stipulations, can be sufficient to promote tolerance and acceptance … Collectively, it appears that contact with a range of viewpoints that differ from one's own can have beneficial effects in a wide range of circumstances. Applying this to the campus setting, more opportunities should be created for all community members to engage with people who think differently from them." (2020, 163)

[61] Christian Picciolini, an author who was recruited by a racist gang at the age of fourteen, explains that isolation plays a role in disgruntled individuals becoming attracted to extremist ideologies: "… the roots of extremist behavior stem from isolation and grievance and form long before there is any focus on ideology. Vital to any effort in reducing the escalating violent extremist threat is the need to help repair the damaged foundations of individuals, instead of shunning them or disparaging them with opposing viewpoints." (2020, 31)

[62] A 1945 quotation from civil rights activist Pauli Murray reads: "I intend to destroy segregation by positive and embracing methods … When my brothers try to draw a circle to exclude me, I shall draw a larger circle to include them. Where they speak out for the privileges of a puny group, I shall shout for the rights of all mankind." (See Lukianoff and Haidt 2018, page 61.)

for employees to receive legal protections that can shield them from punishment by employers for their extramural speech.[63] The goal, of course, is to level the proverbial playing field between employees and employers so that the latter cannot easily censor the former. Even if the policies endorsed by Anderson and Mounk are not equipped to entirely eliminate excessive control of speech by employers, a legislative intervention of this kind could send a powerful signal throughout society about the importance of cultivating social relations wherein people with many different kinds of beliefs and worldviews can dialogue with one another without fear of punishment. Of course, this strategy is relatively demanding since it involves the machinery of the state being deployed in the interest of constraining the ability of employers to meddle with their employees' private lives. Even if the intentions that animate this approach are noble, some will undoubtedly be uncomfortable with the notion that governments ought to be entrusted with greater power with respect to the hiring and firing decisions of private firms.[64]

Some alternative strategies for facilitating inclusion are less demanding in that they simply call upon organizations to voluntarily embrace policies that are conducive to a productive exchange of ideas between people with different perspectives. For example, academic and activist Loretta J. Ross encourages organizations to implement policies regarding conflict resolution that allow people on opposing sides of a dispute to receive a fair hearing, and wherein people with all

[63] Anderson states: "A just workplace constitution should incorporate basic constitutional rights, akin to a bill of rights against employers…A workers' bill of rights could be strengthened by the addition of more robust protections of workers' freedom to engage in off-duty activities, such as exercising their political rights, free speech, and sexual choices." (2017, 68) Mounk similarly argues: "Obviously, employers should be able to restrict what their staff do or say while they are on the job. But unless the nature of their work is openly political … they should not be able to fire their employees for views they express as private citizens. This would go a long way toward giving citizens the confidence to express themselves without fear of material ruin." (2023, 177)

[64] While I do not dismiss the potential value of these kinds of policies, I think that their advocates ought to take seriously that they can backfire. It is possible that if employers are legally barred from sanctioning employees for their extramural speech, this will motivate them to be more selective during the hiring process, and give strong preference to those who share their own views about contentious matters. In such cases, the project of promoting inclusion of diverse perspectives in workplace settings will be undermined rather than strengthened.

kinds of perspectives can express themselves without fear of punishment. She advocates for the adoption of "ground rules" that "make it easy for anyone to speak up—especially against groupthink—without fear of censure, whether they're speaking from a position of power or the margins." (2025, 184) This proposal is clearly aligned with the goal of promoting equity in accountability, as it calls for the most powerful people and the least powerful people within organizations to have opportunities to call one another to account without fear of backlash or reprisal. While creating an organizational atmosphere wherein people with different levels of stature can freely challenge one another is easier said than done, Ross' prescriptions provide a valuable glimpse of what such an arrangement may look like in practice. This is a concrete strategy for fostering inclusion that ought not be dismissed.

An additional approach to fostering inclusion that merits consideration is advocated for by political scientist Verlan Lewis and historian Hyrum Lewis. It is known as "adversarial collaboration", and it involves requiring members of an organization to interact and work alongside colleagues who have been tasked with challenging and critiquing their positions. These authors state:

> The…final step we can take to minimize the scourge of ideology is to engage in 'adversarial collaboration.' This means consciously and systematically incorporating constructive political disagreement into our lives … we can only hope to improve our political understanding by hearing arguments for and against individual positions and evaluating them accordingly. The more we associate only with the like-minded, the more we take our views for granted and the more inflexible and dogmatic we become in those views. Perhaps the best way to check this tendency is to seek out and listen to those who see things differently. Once we have shed the 'one side is right about everything' mentality facilitated by ideological essentialism, then someone who holds a different view on something is not an enemy to be defeated but a partner to be learned from. Such adversarial collaboration has been shown to reduce ideological identification and political error. (2023, 95)

While these scholars do not explicitly cite the work of Mill, their arguments about the value of adversarial collaboration are highly congruent with Millian principles and precepts. They

articulate a vision of human progress, propelled by freedom of expression, that is tacitly Millian if not explicitly so. The authors go on to state: "In science as in politics, we eliminate error and get closer to the truth by subjecting our views to open criticism. Since social progress comes by falsifying incorrect policies and procedures, it also requires an open society that accepts and institutionalizes constructive disagreement." They continue: "Being non-ideological would make us more willing to change our minds in the face of new evidence, which is the key to rationality and, by extension, the key to human progress." (2023, 96) This commentary is apt for our purposes. The notion of institutionalizing constructive disagreement coheres nicely with the arguments about supporting institutions that were articulated earlier in this chapter. It is clearly the case that if institutions become more open to the idea of adversarial collaboration, they will be less inclined to ostracize and eject individuals who question cherished orthodoxies and embrace ideas that are unpopular or controversial. Rather than expelling these people in the interest of cultivating intellectual homogeneity, institutions can channel their energies towards more inclusive and productive ends that can be edifying for all involved. This also means that institutions can challenge wrongheaded or noxious views without engaging in the kind of punitive behaviour that undermines equity in accountability.

While reasonable people can disagree about the wisest way to facilitate inclusion of diverse perspectives throughout various facets of society, the key point to appreciate here is that this project is not merely abstract and theoretical. There are many ways in which individuals and institutions can go about cultivating an atmosphere wherein many different kinds of people, including those who espouse problematic ideas, are given ample opportunity to interact with others in a manner that is conducive to intellectual and social progress. As we have seen, scholars in various disciplines are already taking this matter seriously in hopes of achieving a more open

intellectual climate and a more robust public discourse. Accordingly, there is reason to think that our deeply flawed status quo can give way to a better arrangement over the coming years and decades. The notion that erroneous and offensive views must invariably be met with hostility and exclusion ought to be resisted, as alternative responses are available that are more likely to produce gains for individuals and communities at large.

At this juncture, it is appropriate to return to a point that was made in the introductory chapter. This dissertation does not interrogate the question of whether it is in fact common for people to be targeted with social exile, and to have their reputations and livelihoods destroyed, due to the dynamics of contemporary public discourse. What matters for our purposes is the fear that surrounds this unfortunate fate, and the chilling effects that are generated by this fear. Even if we accept that it is rare for someone to have their reputation and livelihood destroyed as a result of expressing themselves in a way that transgresses social norms, the point still stands that individuals and groups with less economic privilege have much stronger reasons to engage in self-censorship than those with greater economic privilege. This means that social censorship has the potential to erode equity in accountability, in addition to other key social goods such as critical intellectual faculties and authenticity in discourse.

If we wish to preserve and strengthen these social goods, then it is appropriate to work towards the objective of achieving a Millan atmosphere of free expression throughout society, which entails much more than the mere absence of state overreach with respect to expressive acts. The social goods associated with free expression are most effectively cultivated in an atmosphere wherein many different kinds of individuals and groups can explore a vast array of ideas without being stifled by peer pressure and intimidation. An atmosphere of intimidation is incompatible

with an atmosphere of free expression, regardless of whether this intimidation emanates from the state or societal actors outside of the state. The chapter that follows will explore how our contemporary information environment is at odds with the atmosphere of freedom for which Mill advocates.

Chapter iii: How Online Intimidation Culture Undermines Free Expression

iii.i: Social Media and Chilling Effects

The relationship between social media and free expression is currently a hotly contested topic both inside and outside of academia, and it has emerged as one of the central flashpoints in the partisan battles of the 2010s and 2020s. Many researchers and commentators have advanced competing ideas about how much leeway social media users, social media companies, and government regulators should be given when it comes to shaping the discourse that takes place online. Due to the scope and complexity of this subject, not all of these issues can be treated in detail in this discussion. Our particular focus will be on the chilling effects that can be produced by online attacks, and the societal costs associated with these chilling effects. This chapter seeks to make a contribution to debates about social media and free expression by connecting the empirical account of online intimidation culture that was offered in Chapter 1 with the normative, Millian philosophical vision that was explicated in Chapter 2. Our task is to demonstrate that the antagonistic dynamics of social media communication amount to a serious threat to the atmosphere of free expression that Mill and other liberal thinkers wish to cultivate, as well as its associated social goods. This chapter will argue that while the project of developing an atmosphere of free expression and cultivating social goods may on its face appear lofty and distant from the concerns of many ordinary people, there are good reasons to worry about an atmosphere of intimidation making societies less resilient in the face of serious challenges. Accordingly, an atmosphere of free expression ought to be viewed as an asset to society's ability to remain stable over the long term rather than an abstract luxury.

iii.ii: Concrete Cases Involving Social Media Controversy

In order to appreciate social media's ability to limit free expression, it may be worthwhile to briefly examine some examples of social media controversies from recent history that involve social punishment being deployed in response to expressive acts. We can reach a better understanding of online intimidation culture by taking some time to consider how the targeting of individuals via social media can impact public discourse more broadly. As Radzik aptly states: "Informal social punishment has always been a tool of social control. It has always been used both for good and for ill. But social media has amplified the power of informal social punishment considerably. It is time to talk about how this power should be used." (2020, 72) For the purposes of this discussion, it is essential to explore not only how online castigation can affect its direct targets, but also how it can shape the words and behaviour of onlookers who are not directly involved in social media controversies. A key premise that informs this discussion is that social media skirmishes involving relatively small numbers of people can play a significant role in generating outsized chilling effects that shape public discourse in pernicious ways. Even if people who are targets of social media backlash are ultimately able to recover from it, this backlash can still create pressures and incentives that are antithetical to open and productive dialogue.

One noteworthy incident involves the author Kosoko Jackson, who is a high-profile figure in the young adult fiction genre and the community that surrounds it. In 2019, Jackson received harsh blowback in online forums after writing a romance novel set in the context of the Kosovo War in the 1990s.[65] While a few different concerns with the novel were expressed in online reviews

[65] Ilana Redstone and John Villasenor note: "A pair of head-spinning examples of the identity politics–driven meltdown in the world of young adult fiction publishing can be found in cancellation of books by Kosoko Jackson and (separately) Amélie Wen Zhao … In early 2019 Jackson was about to publish his debut novel, called *A Place for Wolves*. The novel was set in 1990s Kosovo and followed the relationship between two American teenage boys.

and social media posts, the most prominent objection concerned its depiction of Muslim people, which was viewed as insensitive. Jackson was criticized for selecting this wartime setting for his novel despite not being a Muslim himself, and for creating a story wherein the protagonists were also non-Muslim US citizens, and members of other demographics were deemphasized. He was accordingly accused of harming members of vulnerable minority populations, and responded to these accusations by apologizing for his actions and calling upon his publisher to pull the book from circulation before its official release date. The publisher assented to his request, and the book was removed from its catalogue.

This case is helpful for understanding the nature of intimidation culture because it involves an agent effectively engaging in self-censorship after becoming a target of online attacks. In this case, the social media users who targeted Jackson had an impact not only on him, but also the sizeable audience that would have chosen to read his novel had they had an opportunity to do so. It is entirely possible that the people criticizing Jackson's book had valid points that merit attention. Perhaps the novel in question truly is deeply flawed, and does a disservice to certain populations by providing an inadequate portrait of a devastating conflict that took place in the real world. We should not dismiss the notion that Jackson's work had problems that deserve to be highlighted. However, what is alarming about this incident is the fact that the online attacks that were launched towards the author put a stop to discourse that may have proven productive had it been permitted to play out in a manner that was less antagonistic and accusatory. Perhaps the author of the book, as well as members of the public who took an interest in it, could have deepened their

Although early signs pointed to the book's potential for success … a lengthy negative review was posted … in February 2019 … This review led to a rapidly mushrooming social media backlash, and then to Jackson himself canceling the book." (2020, 139)

understanding of the Kosovo War and its human toll by taking the time to explore Jackson's text, while also taking seriously the concerns of its critics. Jackson's work, with all of its alleged flaws, could have functioned as an entry point into a valuable and edifying exchange of ideas. Instead, it was removed from the book marketplace, causing dialogue to cease. The fact that Jackson himself called for the book to be removed from circulation only underscores how powerful public castigation can be in causing people to revise their words and behaviour in order to avoid further punishment and remain in the good graces of their community.

Another concerning feature of this episode is the signal that it sends to authors besides Jackson, as well as aspiring authors. Earning a comfortable living as an author is no small feat, and many people who would like to write as a full-time profession are unable to do so. When people are given opportunities to achieve traction and success in the publishing industry, one of the last things that they are likely to do is decide to jeopardize these opportunities by steering towards controversy and harsh criticism. While we can know for sure that Jackson's novel, titled *A Place for Wolves*, was removed from circulation following the accusations that were leveled at him, it is impossible to know for sure just how many other book projects were quietly discarded or abandoned following the publicization of this event. It is entirely possible that authors and publishers alike viewed Jackson's unpleasant experience as a warning sign that if they did not conform to the norms and expectations that pervade the young adult fiction genre, they too would face a social punishment that was as severe or more severe than the one that he had to go through, and potentially end their career in publishing. Indeed, episodes like this one have troubled the community surrounding young adult fiction since social media discourse became ubiquitous,

meaning that it is far from an isolated incident.[66] It is plausible to think that when online castigators successfully get others to pay attention to their condemnations and conform to their demands, this incentivizes further online attacks as social media users learn that they can exert power over authors and publishers if they deploy online platforms in a particular way and appeal to a specific set of prevailing norms and expectations. Even if one does not view the withdrawal of Jackson's book as a major event, it would be an error to overlook its potentially powerful chilling effects and the incentives that it ingrains.

A second noteworthy case that is useful for understanding the dynamics of online intimidation culture involves the data analyst David Shor, who made a social media post in May of 2020 wherein he cited an academic research article by political scientist Omar Wasow. The article argued that while peaceful demonstrations for racial justice had been beneficial for the Democratic Party's vote share in the 1960s, demonstrations that turned into riots were detrimental to the party's electoral prospects. Since Shor had been employed by politically progressive organizations throughout his career, viewers of the social media post could plausibly read him as alerting his followers about the potential for violent protests to undermine the ability of the Democratic Party to emerge victorious from the 2020 US presidential election. In the exceptionally fraught atmosphere of the spring and summer of 2020, wherein discourse about racial justice and injustice in the US became ubiquitous and intense, the backlash towards Shor was swift, and he was fired from his position at the software company Civis Analytics soon after his post went

[66] Adam Szetela offers an overview of such incidents in his 2025 book. He states "On social media, where self-righteous indignation earns more likes and retweets than measured criticism, children's and YA authors are accused of harming and corrupting the next generation. From an African American illustrator honored by the National Association for the Advancement of Colored People to an African American writer and editor who started the first major imprint for African American children's literature, most of the accused are part of the movement for diverse and sensitive books." (2025, 4-5)

online.[67] Fortunately for Shor, he was able to find a new employer within a relatively short period of time, but was prohibited from revealing to the public which firm had hired him.

This case is noteworthy for a few reasons. One remarkable aspect of this episode is the fact that while Shor received backlash for sharing an academic article, little to none of the backlash was directed at rebutting the specific claims offered in this article. Social media users did not appear to express anger at Shor because he presented a view which they believed to be mistaken. Instead, the anger directed at him seemed to emanate from the fact that he was perceived to have expressed an inappropriate attitude towards the enormous groundswell in activism that was taking place at that precise moment. Shor was positioned as a transgressor for reasons that were symbolic, rather than substantive, in nature. The conflict between him and his detractors was not about which party was in possession of the most accurate information or the soundest arguments, but rather about whether Shor had fallen afoul of social norms that were newly ascendant in progressive milieus. While social media users could have interpreted Shor's post more charitably, and read him as making a good-faith effort to provide helpful information to those who care about advancing social justice, he was instead viewed as an antagonist who deserved retaliation, at least to a certain extent. Rather than working alongside one another in order to distinguish between truth and falsehood and forge a sensible path forward in pursuit of common goals, Shor and his detractors became involved in a confrontation over whether the data analyst had expressed sufficient loyalty and commitment to his own ideological camp. While Shor did apologize for his controversial social media post, this did not stop him from losing his job and, at least temporarily, being the

[67] Jonathan Rauch provides a description of this episode: "In 2020 a data analyst lost his job after tweeting an accurate summary of academic research about protest and voting behavior. Twitter users called him a racist, and employees and clients of his company complained that his tweet had threatened … their safety … The analyst apologized the next day … To no avail. He was fired and kicked off a progressive listserv." (2021, 210)

subject of a virtual flogging on social media. His expression was met with social punishment even though he explicitly expressed remorse to those who were angered by it, which further suggests that his transgression and the response to it were largely symbolic.

While it would be an error to suggest that this social media controversy resulted in personal and professional ruin for Shor, as this is not the case, this episode does help illustrate how destructive the dynamics of social media can be with respect to deliberation and debate, even among parties who are broadly in agreement with one another about major goals. It was argued in Chapter 1 that the incentives of social media reward maximalism while sidelining more temperate and nuanced discussion, and this case lends credence to this diagnosis. The academic research that Shor cited on social media suggested that a strategically measured approach to advancing social justice could have more efficacy than a radical approach involving injury to persons and property. This was met with backlash from social media users who were sympathetic towards radical politics, and Shor received a clear signal that his feedback was unwelcome. The issue here is not that Shor's detractors favoured the deployment of extreme measures in order to reach political goals over more moderate alternatives. Rather, the problem is that skepticism towards these measures was shut down via social pressure, when this skepticism could have yielded great value with respect to the overarching objective that Shor and his detractors both shared. It is not just Shor who suffered as a result of his own ideological camp punishing him for his nonconformity, but also the cause of racial justice itself, as people committed to this cause were denied an opportunity to think and communicate constructively about how this cause could best be realized.[68]

[68] Sarita Srivastava points out that a focus on acts committed by individuals can draw attention away from "systemic change and transformation". (2024, 173)

Of course, Shor's loss of employment would only exacerbate the chilling effects associated with this online controversy, as it signalled to onlookers that they too could face severe punishment for questioning the reigning orthodoxy of the moment. Messina offers the following account of why Shor's employer is deserving of reprimand for its response to the online controversy:

> Shor's firing was short-sighted and harmful for the kind of speech environment that we have reason to want. There is reason to believe that it might have increased social pressure within the firm and decreased the firm's diversity in ways that intellectually matter. Accordingly, decision-makers at Civis Analytics should meet frank criticism for their behavior. They have abused the discretion that we have good reason to afford them given the kind of enterprise they run. Their misconduct redounded not only to Shor's detriment but to the detriment of us all. For in caving to a social pressure campaign, and in making casual speech the grounds for dismissal, the firm's leadership has helped sustain an environment in which many more will refrain from expressing themselves to avoid meeting with similar fates. (2023, 85)

It will come as no surprise that I share Messina's concerns about chilling effects in this case. However, legitimate concerns about the pernicious impact of this episode do not end there. Perhaps even more alarmingly, the expulsion of Shor sent a signal that if people wanted to publicly question the ideas and behaviour of hardline activists, they would need to do so in a venue that was explicitly divorced from progressive politics. If Shor had decided to immerse himself in a milieu that was clearly divorced from progressive politics, it is very possible that he would have received no punishment for his social media post, or would instead have been rewarded for expressing concerns about rioting. By censuring a member of their ideological camp and prompting his firing, the social media users castigating Shor indicated to onlookers that critical feedback would be met with hostility, thereby incentivizing these onlookers to voice their concerns in alternative settings wherein they would be safe from social punishment. Without necessarily intending to do so, the people attacking Shor helped to facilitate siloing of the media ecosystem, and society more broadly, by promoting the idea that critical feedback in progressive spaces would

lead to ejection. This kind of behaviour clearly undermines intellectual diversity, and can even play a role in hardening antagonism between people with different views by preventing them from having conversations that might generate greater understanding and tolerance. This incident demonstrates the power of social media to function as an engine of division in addition to conformity, which of course is profoundly injurious to productive discourse.

A third social media controversy that merits consideration involves a Palestinian-American family that had established a successful grocery business in Minneapolis, Minnesota. In this case, old social media posts from the business owner's daughter were brought to light that clearly contained bigoted and offensive material. Despite the fact that this individual was a teenager when the posts were created, social media users quickly organized and called for a boycott of the family's business. The business owner, Majdi Wadi, responded to the intense backlash by firing his daughter from the company, but this did not successfully deter activists who ostensibly wanted to see the enterprise eliminated from the marketplace on the grounds that the people affiliated with it were guilty of promoting, or at least tolerating, racism. The boycotts led to business partners cancelling ties with the family, resulting in the loss of millions of dollars in expected earnings. Accordingly, nearly seventy employees were laid off as the family struggled to manage the turmoil that had been inflicted upon their company. While the Wadi family's business remains in operation, it is clear that that the organized backlash was costly to them both personally and financially.[69]

[69] Journalist Robby Soave writes: "Consider Majdi Wadi, a Palestinian immigrant to the United States who operated a catering business in Minneapolis that employed two hundred people … Wadi fired his daughter in order to save his business, but it didn't work: All of his business partners canceled their contracts, and his landlord terminated his lease." (2021, 170-171)

This social media controversy has some similarities with the case involving Shor due to the context in which it took place and the nature of the accusations leveled at the Wadi family. However, it is different from the aforementioned case because the social media posts that led to social punishment had been created years prior to the controversy taking place, and because it involves targets of social punishment who are separate from the actual offender. Even if we accept the premise that Majdi Wadi's daughter was an appropriate target of some form of punishment for her bigoted and offensive social media posts, it is not clear why this punishment should extend to her family members, who had no involvement in the posts. This case also raises questions about whether it is appropriate for adults to receive punishment for bad acts that took place before they reached adulthood, although these concerns are less relevant to the overarching goals of this discussion. What is primarily of interest here is the fact that social media can not only be deployed in order to administer social punishment to people who fall afoul of social norms and mores, but also the people who are affiliated with them.

This social media controversy raises serious concerns about the ability of online attacks to constrain the freedom of people who are not directly involved in any sort of offensive expression or behaviour. If it is the case that people can be subjected to severe social punishment not only for their own words and actions, but also those of people with which they have social ties, then this will significantly impair people's ability to freely form associations and relationships with people who are different from themselves. If any individual can be targeted with social punishment simply because they bear some relationship with a person who has committed a transgression, even when this transgression lies years in the past, then this dynamic will chill not only discourse, but the process of socialization that is a prerequisite for discourse. In this kind of atmosphere, people will

need to be extremely cautious and selective when it comes to their social ties in order to shield themselves from the pain that online attacks can bring. Any mistake with respect to the process of forging relationships with others will carry with it the possibility of castigation and accusations of guilt by association, which is a frightening prospect for most people. The notion that episodes like the one involving the Wadi family simply include accountability for poor conduct are clearly flawed when we consider how many people were adversely affected by the poor decisions of a single young person. It is clear that episodes like this one have the power to establish an atmosphere of intimidation wherein all kinds of people must exercise great caution with respect to their own words and conduct, as well as the social ties that they decide to establish. Such episodes can be profoundly corrosive to public discourse and the worthwhile goals that can be achieved through good-faith communication between diverse individuals and groups.

iii.iii: The Scope of Intimidation Culture

Earlier portions of this discussion have argued that social media platforms function as engines of conformity, and that this conformity is achieved through intimidation. People – especially those with moderate views – often abstain from saying what they really think in the realm of social media because doing so can involve significant social penalties. At this juncture, it is appropriate to say more about the scope of intimidation culture, and substantiate the notion that attacks that are launched in online spaces can really shape society in meaningful ways. If we are going to make the claim that social media undermines an atmosphere of free expression and supports an atmosphere of intimidation, then we must consider the interplay between online content and offline expression and behaviour. Some may scoff at the notion that words and images that are circulated online can generate chilling effects that transcend social media platforms and

implicate a broad array of societal actors. However, I will endeavour to demonstrate that there is little reason to think that the impact of these dynamics is confined to the online realm.[70]

Bail explains how the dynamics of social media generally amplify the voices of those with extreme views while chilling the expression of those who are more moderate in character. He states that in many cases, people in this latter category fear that online interactions will bring harm to them in offline settings:

> Extremists … turn to social media because it provides them with a sense of status that they lack in their everyday lives, however artificial such status might be. But for moderates … the opposite is often true. Posting online about politics simply carries more risk than it's worth. Such moderates are keenly aware that what happens online can have important consequences off-line. (2021, 77)

After interviewing an array of social media users with varying political orientations, Bail notes that "[i]n addition to having concerns about their livelihoods, many of the moderates … interviewed were worried that discussing politics on social media would upset their family members or friends." (78) These comments lend strong support for the idea that the toxicity that pervades online discourse can generate chilling effects that transcend the online realm.

Bail goes on to compellingly argue that the dynamics of social media lead to a distorted view of society as hardcore partisans continuously capture widespread attention while sidelining and marginalizing their more measured counterparts. He also notes that these dynamics are likely to drive antagonism and estrangement between different segments of society:

> What does it mean that moderates are missing from social media discussions about politics? In my view, this is the most profound form of distortion created by the social media prism … the social media prism makes the other side appear monolithic, unflinching, and unreasonable. While extremists captivate our attention, moderates can seem all but

[70] Stephen Macedo and Frances Lee highlight social media as a factor that fueled intolerance during the pandemic era. (2025, 284)

> invisible. Moderates disengage from politics on social media for several different reasons. Some do so after they are attacked by extremists. Others are so appalled by the breakdown in civility that they see little point to wading into the fray. Still others disengage because they worry that posting about politics might sacrifice the hard-fought status they've achieved in their off-line lives. Challenging extremists can come back to haunt moderates, disrupting their livelihoods, friendships, or relationships with family members they will see every year at Thanksgiving … Unfortunately … opportunities for mutual understanding are few and far between in an age of rapidly increasing social isolation. As Republicans and Democrats continue to sort themselves into separate ZIP codes, pastimes, and social circles, I worry that the power of the social media prism to fuel extremism and mute moderates will only continue to grow. (2021, 82-83)

Research suggests that self-censorship is now a common feature of everyday life that transcends online settings,[71] and the fact that general reports of such silencing increased dramatically over the course of the 2010s is consistent with the view that social media plays a role in producing self-censorship that spills over into the offline domain.[72] Academia, journalism, and publishing are salient examples of professional domains wherein intimidation on social media reportedly shapes offline behaviour, including the expression of ideas.[73] While it is certainly true that offline self-censorship can have causes that do not specifically involve social media, it would be naïve to think that this technology does not play a role in this area, as it has become thoroughly intertwined with virtually all facets of modern life. There is room for reasonable debate about the extent to which offline self-censorship is fueled by social media dynamics, and it is simultaneously

[71] Mounk states: "Incidents of censorship or social shunning attract most attention when they involve someone famous. But in the main, they affect ordinary people who never make the news. More than three out of five Americans now say that they abstain from expressing their political views for fear of suffering significant adverse consequences. A majority of college students report having self-censored in the past, with only one out of every four saying that they are comfortable discussing controversial topics with their classmates. Even at *The New York Times*, about half of the paper's own employees believe that many of their colleagues are 'afraid to say what they really think.' (2023, 164)

[72] See Figure 1 and Figure 2 in Gibson and Sutherland (2023).

[73] FIRE president Greg Lukianoff and Jonathan Haidt have the following to say about the "transformative" effect of social media: "Call-out culture requires an easy way to reach an audience that can award status to people who shame or punish alleged offenders. This is one reason social media has been so transformative: there is always an audience eager to watch people being shamed, particularly when it is so easy for spectators to join in and pile on." (2018, 71-72) They offer the following insights: "Life in a call-out culture requires constant vigilance, fear, and self-censorship. Many in the audience may feel sympathy for the person being shamed but are afraid to speak up … Reports from around the country are remarkably similar: students at many colleges today are walking on eggshells, afraid of saying the wrong thing, liking the wrong post, or coming to the defense of someone whom they know to be innocent, out of fear that they themselves will be called out by a mob on social media." (2018, 72)

reasonable to posit that social media likely plays a role in this self-censorship given the ubiquity of this technology, as well as the fact that self-censorship has increased in general as social media platforms have grown in popularity and influence.

Redstone and Villasenor draw particular attention to the realm of academia and explain how social media toxicity can shape the behaviour of people who have no interest in participating in online conflicts:

> …social media can act both directly (e.g., through call-out campaigns) and indirectly (through behavior modification aimed at avoiding social media opprobrium) to shape what happens on campus. One reason is that technology has upended how everyone—including the academic researchers we entrust to discover and disseminate new knowledge and the professors and other instructors we entrust to teach college classes—communicates. Another is because academia and the pursuit of knowledge have always been closely linked to broader political, social, and religious currents. Social media provide a new feedback mechanism through which those currents can shape and be shaped by what happens on campus … public shaming on platforms such as Twitter and Facebook is an extraordinarily effective tool for behavior modification. Individuals who have been targeted by call-out campaigns highlighting real or perceived transgressions will be less likely to do anything in the future that might once again attract online wrath. Even people who have not been targets of call-out campaigns see what happens to those who have, and will modify their behavior as well to avoid becoming targets themselves. (2020, 43)

Social media's power to chill discourse in the realm of academia is concerning, as academic institutions are explicitly committed to the pursuit of truth and the dissemination of knowledge. Accordingly, academic institutions will be derelict in their duties if they establish a culture of conformity wherein orthodoxies are protected from scrutiny and dissenting views are suppressed.

Unfortunately, similar reports of the chilling power of social media have emerged from the world of journalism. *Newsweek* deputy opinion editor Batya Ungar-Sargon has the following to say about the influence of online intimidation with respect to journalists and the reporting and commentary that they provide to the public:

> …the thing is, you don't actually have to weed out every heretic for public shaming to be effective at silencing dissent; after a while, people silence themselves. Who would volunteer to go through that kind of bullying, when they could avoid it by staying quiet? The spectacle it creates on its own has a powerful effect on enforcing compliance, creating a public sphere in which an angry online mob has more power to silence journalists, through peer pressure, than do the editors of the most important news organizations in the world. (2021, 172)

The publishing industry similarly appears to be affected by conformist pressures that seek to block various perspectives from receiving a hearing. FIRE president Greg Lukianoff and columnist Rikki Schlott argue that individuals involved in the publishing industry have increasingly embraced a censorious attitude over the course of recent years:

> Over the past several years, the publishing and literary world has been consumed by cancellations aimed at staffers and authors, high- and low-profile individuals alike. We depend on the publishing industry to proliferate ideas and act as a viewpoint-neutral platform for a wide host of authors and thinkers to share their thoughts … A new generation of employees in the publishing world seem exceptionally comfortable assuming the role of ideological gatekeepers—and have a hard time distinguishing books that might not be 'their cup of tea' from those their publisher should abandon. (235-236)

Moreover, Lukianoff and Schlott explicitly draw attention to social media as a factor in intimidating authors and causing certain views to be omitted from public discourse: "As it turns out, the notion that a mounting Twitter mob might turn on you strikes so much fear into the hearts of some authors that they move to proactively censor themselves, or even cancel themselves. And not even the sensitivity readers—who are hired to look for offensive content in other people's books—are safe from the mob." (2023, 242)

Censoriousness in the publishing industry is the focal point of Adam Szetela's book *That Book Is Dangerous!: How Moral Panic, Social Media, and the Culture Wars Are Remaking Publishing.* This author too draws attention to the ability of online platforms to facilitate intimidation and conformity in the domain of publishing:

> … the movement for more diverse and sensitive books has created new problems. In the past decade and a half, the emergence of platforms as different as Twitter, Tumblr, TikTok, and Goodreads has allowed anyone with an internet connection to be a public literary critic. … Far from irrelevant, this movement uses social media and other platforms to pressure authors, agents, and editors to abandon the idea that people should be allowed to write and read what they want. (2025, 4)

All of these contributions lead to the conclusion that intimidation on social media can have very real consequences in an array of offline settings. This is something that ought to be confronted by those who are skeptical towards the idea that social media is a serious threat to the free exchange of ideas.[74] A crucial observation is that the evolution of technology has made it increasingly difficult to make sharp distinctions between the offline domain and the online domain.[75] The ubiquity of smartphones has made it easy for users to capture audio and video of others and share it via social media, regardless of whether these individuals wish to receive online attention, meaning that moments that were intended to remain private can be made public with very little in the way of thought and consideration.[76] It is also the case that social media posts criticizing the words and actions of an individual can receive significant traction regardless of whether they include a fair characterization of what the individual said or did, meaning that in many cases, individuals have reason to take proactive measures to prevent misunderstanding and intentional distortion in the realm of social media.

---

[74] Adrian Daub criticizes widespread concerns about cancel culture: "My suspicion is that complaints about cancel culture don't really solve anything, nor are they meant to. They are rather part of a moral panic. This is primarily related to … the attention economy: People talk about cancel culture so that they don't have to talk about other things, in order to legitimize certain topics, positions, and authorities and delegitimize others." (2024, 1)

[75] Alice E Marwick states: "Bullying is social and often enabled by technology… bullies take advantage of social media's leakiness, which enables context to be stripped and a false context supplied so that posts are misunderstood." (2023, 144-145)

[76] Francis Fukuyama addresses the issue of private utterances being circulated to online audiences: "Private views that previously would have been expressed in person or over the telephone are now mediated by electronic platforms, where they leave a permanent record … Many users express what they believe are private views through email or to small groups of people on social media. Anyone receiving the message, however, can broadcast it to the rest of the world, and many people have gotten into trouble in recent years simply for speaking honestly in what they believed to be a private setting. There is, moreover, no statute of limitations on the internet; anything you say becomes part of a permanent public record that is extremely difficult to disavow subsequently." (2022, 107)

Even when people are not directly participating in social media discourse, it makes sense for them to remain guarded in their interactions with others in order to avoid online castigation. In some cases, the people who are targets of social media vitriol have little or no interest in being a participant in online battles, but find themselves in such a situation regardless. It is possible that we are only in the early stages of understanding just how powerful social media has been in generating self-censorship in offline settings, even among those who have little affinity for social media in general. While it would be an overstatement to say that it is impossible today for people to find refuge from the intimidation and conformity that pervade social media, there are good reasons to worry about this intimidation and conformity seeping into areas of life that extend far beyond the online realm.

iii.iv: An Atmosphere of Intimidation

The above commentary leads to the conclusion that social media has the power to establish an atmosphere that is antithetical to the one that Mill and other liberal advocates for free expression wish to cultivate. While the kinds of punishment that are doled out via social media are obviously not equivalent to formal, state censorship, they can and do amount to a form of social censorship that seriously undermines intellectual and expressive freedom. As we saw in the previous chapter, Mill's 1859 text *On Liberty* argues that social censorship is even more dangerous than political censorship due to its ability to influence virtually all facets of human life and have a suffocating effect on people who wish to think independently and explore ideas without conformist pressure permeating around them. Mill's arguments seem even more apt in an age when people can receive attention and notoriety at any moment thanks to the speed and distance at which information can

travel with little effort. If the prospect of being targeted with vitriol and condemnation by members of one's social circle is frightening for people, then the prospect of becoming a target of opprobrium for thousands or millions of strangers may plausibly be viewed as downright terrifying.[77] Now that people equipped with smartphones can not only share text across the globe with ease, but also high-definition pictures and videos (including livestreams),[78] people have more opportunity to receive unwanted attention from others than ever before, and this can include vicious attacks that go well beyond the domain of good-faith criticism. While freedom of expression is obviously our overarching concern in this discussion, it is worth noting that these dynamics also raise concerns about privacy[79] that overlap with concerns about intellectual and expressive freedom, and these ought not be overlooked.

Interestingly, Mill argues that social tyranny over individuals has a tendency to grow more intense over time. Let us consider the following passage:

> Apart from the peculiar tenets of individual thinkers, there is also in the world at large an increasing inclination to stretch unduly the powers of society over the individual, both by the force of opinion and even by that of legislation; and as the tendency of all the changes taking place in the world is to strengthen society, and diminish the power of the individual, this encroachment is not one of the evils which tend spontaneously to disappear, but, on the contrary, to grow more and more formidable. The disposition of mankind, whether as rulers or as fellow-citizens, to impose their own opinions and inclinations as a rule of conduct on others, is so energetically supported by some of the best and by some of the worst feelings incident to human nature, that it is hardly ever kept under restraint by anything but want of power; and as the power is not declining, but growing, unless a strong

[77] Mari J. Matsuda notes the effect of hate propaganda its targets: "As much as one may try to resist a piece of hate propaganda, the effect on one's self-esteem and sense of personal security is devastating. To be hated, despised, and alone is the ultimate fear of all human beings." (2018, 25)

[78] DiResta highlights the power of livestreams to intimidate: "… what happens online sometimes spills into the real world in frightening ways. … threats and intimidation aren't criticism and oversight. Most people would find being recorded or photographed by a stranger as they walked down the street disconcerting and threatening; having the moment livestreamed for a mob makes it worse. (2024, 283)

[79] Laura DeNardis maintains that the ubiquity of internet-connected devices lays the groundwork for injury to privacy: "… data gathering of routine activities within homes and around medical and health practices can be much more privacy invasive even than surveillance of emails, texts, websites visited, and other digital content through the clear portal of a screen." (2020, 4)

> barrier of moral conviction can be raised against the mischief, we must expect, in the present circumstances of the world, to see it increase. (2015, 16-17)

If we accept the Millian idea that society's conformist pressures tend to grow "more formidable" over time, then it is plausible to view technology, and social media in particular, as an accelerant that makes it even more difficult for individuals to find refuge from the intense judgment that accompanies society's power over the individual. It is arguably the case that the dynamics of social tyranny are now less constrained than ever before thanks to the ability of the Internet to make information available at any time, all across the globe, to enormous audiences. Throughout most of history, if a person had their reputation sullied in a specific geographical location, they could cope with this problem by relocating to another area in hopes of starting afresh. The power of society over the individual was limited by simple logistics, and the fact that in many cases it was impossible for particular communities and societies to transmit damaging information about an individual across vast portions of the Earth's surface. Relocation as a strategy for coping with reputational damage is much less promising in a modern era wherein information can travel around the world with minimal effort.[80] When we take time to reflect on how influential social media platforms, in addition to search engines such as Google, can be in impacting people's social standing and the opportunities available to them, Mill's complaint about the progressive "encroachment" of society over the individual seems prescient.

It was argued earlier in this chapter that online controversies involving relatively small numbers of people can generate significant chilling effects. At this point we should have a deeper

[80] Radzik point out: "Courts of law also allow the accused to defend themselves before inflicting punishment. Twitter does not. The ones punished may be unable to broadcast their defenses as broadly as their shame was broadcast. Sober corrections of the record garner far fewer likes and reposts than juicy accusations and witty denunciations. Even if the accused can make themselves heard, the damage may already have been done." (2020, 52-53)

understanding of why this is the case. Social media makes it possible for individuals and groups to create digital content that can be viewed by virtually anyone on Earth with Internet access. When social media users hurl serious accusations at people, these accusations may reach a broad audience and seriously harm their target's reputation or livelihood. Even if the likelihood of reputational and professional ruin is low in any particular case, the great power of the Internet presents the opportunity for damaging content to have incredible reach both geographically and temporally, which can be highly intimidating.[81] When considering these dynamics, it is important to bear in mind that in many cases, the potential gains associated with challenging orthodoxies and entrenched patterns of behaviour are modest. While it is sometimes the case that dissident and contrarian voices are applauded and rewarded for their contributions to discourse, in many other cases they are simply met with irritation and hostility. For many people, the potential downsides of challenging prevailing ideas and practices will overwhelm the potential upsides, and they are likely to simply refrain from expressing unpopular views for the sake of blending in with their broader community and avoiding the pain of castigation and exclusion.

All of these observations point to the conclusion that online intimidation culture is more than an irritant or a distraction from more pressing issues.[82] It is a force that has the potential to seriously undermines people's intellectual and expressive freedom, and the broader atmosphere of free expression that Mill and others wish to cultivate throughout society. Where an atmosphere of

[81] Content that is posted online can continue to be accessible for very long periods of time, with no sign as to if and when it will be made inaccessible. In the Wadi family case, the bigoted and offensive social media posts that led to the family's turmoil resurfaced after having been deleted.

[82] Radzik states " … naming and shaming through social media looks like a particularly problematic method for informal social punishment. It is especially vulnerable to objections about determining guilt, punishing proportionately, avoiding unintended consequences, ulterior motives, and effectiveness in moving wrongdoers toward atonement and reintegration into in the community. Perhaps a good provisional rule is that informal social punishments should be delivered privately unless there is a good reason to punish publicly." (2020, 59)

intimidation exists, an atmosphere of free expression cannot flourish. The two are incompatible, and if it is true that technological advancement has the ability to augment society's power over the individual, then it is incumbent upon us to develop a thorough understanding of social media's pernicious impact on public discourse before even greater threats emerge in the technological and media landscape. If we are partial to Mill's view that intellectual freedom and expressive freedom are inextricably linked, then we must entertain the possibility that at a certain point, living in an atmosphere of intimidation will begin constraining people's thinking as well as their public speech. Left unchecked, an atmosphere of intimidation may produce a status quo wherein heterodox thoughts generate so much discomfort for individuals that they will simply be suppressed, because individuals know that letting these ideas develop will make them vulnerable to significant pain.

It is worthwhile to note that there is a mutually reinforcing relationship between an atmosphere of free expression, and institutions that support free expression. It was argued in Chapter 2 that in a Millian framework, an adequate atmosphere of free expression may require institutions to be designed with the goal of equipping populations with skills related to argumentation and debate, as well as creating spaces wherein dialogue between diverse individuals and groups can unfold. Indeed, it is reasonable to think that supporting institutions can play a key role in establishing a "strong barrier of moral conviction", to use Mill's phrase, against the power of mobs to stifle individuality and the liberties that it entails. The point that I wish to make here is that while it is true that well-designed supporting institutions can make it easier for an atmosphere of free expression to thrive, it is also true that an atmosphere of free expression can make it easier to build and reform institutions in productive ways. This is a bidirectional relationship rather than a unidirectional one. Institutions are never static, and constantly require a certain amount of

revision to their policies and practices. The more that a society rejects an atmosphere of intimidation and embraces an atmosphere of free expression, the better equipped it will be to maintain institutions that are conducive to quality public discourse. The more that a society succumbs to an atmosphere of intimidation, the more difficult it will become to modify institutions in productive ways, as these institutions will be primed to reject and even punish individuals and groups that critique them. One of the dangers associated with an atmosphere of intimidation is that such an atmosphere can perpetuate itself by blocking criticism and reform efforts that could have a liberatory effect on humans' intellects and expression.

iii.v: Damage to Social Goods

Earlier portions of this dissertation have argued that the dynamics of social media tend to reward extremity. Since social media platforms amplify content that is most likely to receive engagement from audiences, this means that users are incentivized to communicate and present themselves in a manner that is maximally attention-grabbing. Social media companies are much more concerned with the quantity of discourse that takes place on their platform than the quality, as keeping consumers viewing and engaging with social media interfaces is ultimately what is most beneficial to these companies' financial interests. These incentives mean that social media platforms can have a caricaturing effect on public discourse, as they reward participants for presenting their own views in exaggerated ways, and also for engaging in exaggerated antagonism with others who express different views. The drama of interpersonal social media battles is highly enticing to online audiences, and can provide a powerful avenue for increasing one's own public profile and building a sizeable following.[83] This means that the incentives of social media

[83] Daniel F. Stone addresses how greater engagement leads to greater circulation and prominence: "We're more likely to see content from our network that's been 'liked' and shared more often, which is disproportionately likely to flatter

discourse are consistently arrayed against users who are interested in having discussions that are measured, nuanced, and charitable, as participants of this kind are often drowned out by others[84] who do a better job of appealing to audiences' emotions and ascending the social media ranks.[85]

Having argued that these dynamics, and their associated chilling effects, make social media discourse a serious threat to a Millian atmosphere of free expression, at this point it is appropriate to reflect on the three social goods that were explored in detail in the previous chapter. It is worth taking a moment to consider how these specific social goods are impacted by the extremity of online discourse and the intimidation that it facilitates. Critical intellectual faculties, authenticity in discourse, and equity in accountability are three social goods associated with free expression that I have chosen to explore in detail in this discussion, and we will examine how the incentives that are present in contemporary online discourse intersect with these important goods. I will argue that the dynamics of social media discourse have a corrosive impact on these goods, and accordingly make society worse off than it otherwise could be.

---

our side and pillory the opposition, especially when expressed with 'moral-emotional' language … Out-party hostility drives engagement on social media … and is the primary motivation behind sharing fake news in particular …The fact that posts and tweets loudly expressing anger toward the out-party are more likely to go viral can incentivize strategic outrage and distortion for users trying (perhaps unconsciously) to maximize engagement, making (false) outrage-infused content even more common." (2023, 126)

[84] Stone explains how more extreme behaviour on social media overshadows less extreme behaviour: "... Active social media users tend to be relatively extreme, close-minded, overconfident—and more affectively polarized. Moreover, when more typical partisans are politically active online, we can act in a way that is not typical of ourselves. Sometimes we're more disrespectful, belligerent, and aggressive—and get more attention when we act this way. And even when we aren't trying to be combative, we dehumanize our online interlocutors and are relatively likely to be interpreted uncharitably by others." (2023, 125-126)

[85] Srivastava states that social media escalates the emotional nature of contentious debates: "The speed and anonymity of social media has even further heightened the emotionality of debates about race and diversity. This emotional milieu in turn supports the targeting of racialized people. One example is the simmering resentment and anger that has fueled anti-immigrant campaigns around the globe." (2024, 28-29)

Following Mill, I endorse the view that critical intellectual faculties are best developed in environments wherein people have the opportunity to examine issues from many possible perspectives. Instead of simply reaching a firm conclusion about an issue and sticking to it, people can better hone their critical thinking skills by examining the strongest possible case that can be made in support of all kinds of conclusions, including ones that are clearly erroneous. It is only by comparing and contrasting diverse perspectives on a single issue that people can deepen their understanding of it, and arrive at conclusions that are truly informed and balanced. Of course, in order to actually do this, people must have access to venues for expression wherein they will not be punished for entertaining ideas that are wrong, unusual, offensive, or otherwise objectionable. People must have the freedom to explore ideas in an unrestricted manner, and to challenge prevailing orthodoxies, regardless of how commonsensical they may seem. An atmosphere of free expression enables people to grapple with challenging ideas and sharpen their critical intellectual faculties to the greatest extent possible, which in turn helps them enhance their own pursuit of truth, as well as the broader societal pursuit of truth in which they are immersed.

Unfortunately, the dynamics of contemporary online discourse undermine this process by putting enormous pressure on people to fit neatly into ideological camps.[86] It is vanishingly rare to find examples on social media of people taking seriously the views of their intellectual adversaries and seeking to reconstruct them in a manner that is accurate and charitable. While this

[86] Tosi and Warmke point out how off-putting social media can be: "Many people have little tolerance for constant displays of anger. The whole business is unpleasant, and few of us would ever want to be the target of an online shaming mob." (2020, 88) They go on to state: "It is bad for everyone when moderates check out of public moral and political discourse. The most obvious negative effect is that the people who avoid such discussions don't hear arguments and evidence for other views, so their own beliefs go untested. It's easier to maintain your poorly formed convictions if you never discuss them with others, who might show that you're mistaken. But perhaps even worse, when people keep their beliefs to themselves, the rest of the world is deprived of thoughts they otherwise might never encounter … A healthy public discourse takes all kinds. So when the domain of actively discussed ideas shrinks, we are all worse off for it." (2020, 90)

is precisely the type of engagement that is taught and encouraged in Philosophy classrooms, it is a type of engagement that is rendered nearly impossible in our modern media ecosystem. Instead of being rewarded for engaging with intellectual adversaries in a manner that is fair and charitable, social media users are likely to be punished for doing so, as this can signal insufficient loyalty to one's own ideological camp. One does not "win" the game of social media, garnering likes, comments, and followers, by carefully and exhaustively explicating a variety of rival positions and explaining which among them is the most sound. Rather, one "wins" in the realm of social media by loudly proclaiming their allegiance with a specific group[87] and launching attacks on those who fall outside of it. If social media users castigate their opponents with ad hominem attacks, strawmen, and accusations of guilt by association, this can generate even greater rewards by signalling one's passionate dedication to their own ideological camp and its associated orthodoxies. The dynamics of online discourse create an environment wherein people are encouraged not to explore a broad range of views and then carefully reach their own conclusions, but to attach themselves to a larger group, and then increase their social standing within this group by loudly promoting its slogans and talking points. These incentives are destructive to critical intellectual faculties and the project of developing them over time.

If it is true, as I have argued, that the chilling effects associated with online discourse have the potential to constrain expression in offline domains, then these concerns about critical intellectual faculties being stunted cannot simply be dismissed on the grounds that social media platforms are venues that attract people who care little for critical intellectual faculties in the first place. Even if this notion is true, we still must contend with the fact these social media users have

[87] Joshi highlights how the dynamics of social media engagement can be conducive to status-seeking. (2021, 84)

the ability to influence public discourse at large in ways that are clearly pernicious. The aggressive behaviour of social media users has the potential to strike fear in people involved in countless domains, including ones that are explicitly connected to the cultivation of critical intellectual faculties and the pursuit of truth. People in academia, journalism, publishing, and other knowledge-generating domains can be intimidated by social media discourse, and accordingly modify their own speech and behaviour in order to avoid online backlash. It is thus reasonable to conclude that the dynamics of social media discourse are not only damaging to the critical intellectual faculties of people directly involved in these online venues, but also to the critical intellectual faculties of people operating in offline venues who would simply like to explore ideas in a freewheeling fashion without becoming a target of online attacks. A relatively small number of zealots can inflict major harm to the social good of critical intellectual faculties thanks to social media and its unprecedented reach into all facets of society.

These comments are equally applicable with respect to the social good of authenticity in discourse. If people must modify their words and behaviour, perhaps even going so far as to engage in preference falsification, in order to avoid aggression from social media users, then this is obviously injurious to authenticity. If a person is tailoring their words and actions in order to avoid social punishment, then they are not engaging with others in a manner that is genuine.[88] This is true even when the social punishment is entirely hypothetical and the people who are in a position to administer it are strangers. As we have seen, the ubiquity of social media means that people who do not directly participate in online discussions can be pressured into concealing their genuine

[88] Joshi states: "… we're often tempted to seek status at the expense of doing good work, or seek the pleasure that comes with social praise at the expense of being authentic … temptation and self-deception are part and parcel of human life." (2021, 96)

views for the sake of avoiding social punishment. While it is impossible to quantify exactly how much inauthenticity has been generated by the rise of social media, it is clear that this technology has introduced incentives into public discourse that direct actors away from the project of communicating with others in pursuit of the truth, and towards the project of displaying their allegiance to prevailing ideas for the sake of garnering the approval of others.[89]

The architecture of social media can also be damaging to authenticity in discourse in more insidious ways. Anyone who has experience with social media knows that the success of an account is largely dependent on its ability to take advantage of features of the social media interface such as profile pictures, profile banners, profile biographies, thumbnails, titles, hashtags, captions, and the like. A social media user who knows how to create visuals that are highly enticing to the human eye will find much more success than a user who struggles in this area. Social media content that dazzles the senses is much more likely to achieve traction and amplification than content that is less emotionally arousing. What this means in the aggregate is that the architecture of social media is largely hostile to nuance. While particular social media users may in fact have ideas, tastes, and aspirations that are fairly sophisticated, the medium itself puts pressure on these people to present a simplified version of themselves in order to receive attention and validation in online spaces. Social media can undermine authenticity by incentivizing users to "dumb down" their authentic selves for the sake of constructing an online avatar that is more palatable for online audiences that have access to a staggering amount of content at any given moment. Since the

[89] Rose-Stockwell insightfully notes: "In this competition to gain the approval of the audience, grandstanders often make up moral charges, pile on in cases of public shaming, and state that anyone opposing them is obviously wrong … Because of the sheer number of observers and our tendency to seek signals from our online communities, many of our disagreements on social media become metrics-driven opportunities for grandstanding … When others are ranking and scoring us in real time, we lose the ability to examine new concepts in good faith." (2023, 141)

environment of social media is highly competitive, users can experience losses with respect to their own popularity and reach if they fail to give audiences more of what they want, and it is rarely the case that these audiences are seeking complexity, detail, and nuance.[90]

While it is true that this phenomenon can involve people presenting more extreme versions of themselves in online settings with respect to their political views, thereby laying the groundwork for perpetual conflict with others, these dynamics extend far beyond the realm of politics and the culture wars that are intertwined with politics. Social media rewards extremity with respect to many types of content. Social media users who achieve traction by engaging in eccentric humour are incentivized to continue getting more eccentric over time. Social media users who achieve traction by posting sexually provocative content are incentivized to continue getting more provocative over time. Social media users who achieve traction by engaging in dangerous thrill-seeking activities are incentivized to continue getting more dangerous over time. The point is that the incentives of social media reward people for providing audiences with content that adheres to a consistent tone and aesthetic, rather than for presenting themselves as complex individuals with an array of diverse characteristics. While we have seen that social media can fuel intimidation by amplifying aggressive speech and downplaying speech that is more measured, it is important to recognize that social media can also fuel intimidation by signalling to people that in order to receive attention and affirmation, they must present themselves in a manner that is one-dimensional and easy to market to large audiences. In a world that is increasingly shaped by social

[90] Emily Hund similarly writes: "Influencers readily acknowledge that, despite their appearances of being forthcoming, the personal brand is obfuscatory by necessity. Individual personalities are too complicated and contradictory to be captured in the clear, bullet-point legibility required by advertisers, so a distancing occurs: *this is me, and this is my personal brand*." (2023, 42-43)

media, resisting the pressure to present one's self in an oversimplified manner can require significant resolve.

The upshot of all of this is that even when we look beyond the realm of politics and its associated culture wars, the incentives of social media are generally arrayed against authenticity in discourse. Because of the fact that social media users achieve success by cultivating an audience and catering to its expectations, they face risks anytime that they upset these expectations. Social media rewards its users for being predictable and consistent rather than for being authentic. While authentic public discourse involves a certain amount of vulnerability, and willingness to confront complexity and the uncertainty that comes with it, these traits are inconsistent with the manner in which social media platforms function. Indeed, new concepts and phrases are now emerging that help bring these dynamics into clear view. The concept of "audience capture" has begun to enter the mainstream in recent years,[91] and it denotes the phenomenon of individuals and organizations becoming excessively influenced by the feedback that they receive from audiences, which can cause them to drift away from their core convictions and objectives for the sake of maintaining popularity. Audience capture occurs when actors become aware that they may suffer losses as a result of challenging their own followers, and choose to cater to said followers in order to continue receiving engagement, approval, and the financial benefits that these can bring. It is impossible to know for sure how many high-profile individuals and groups have betrayed their convictions as a result of social media dynamics. However, it is certainly the case that the pressures of social media

[91] In his discussion of audience capture, Rose-Stockwell states: "We respond to the types of positive signals we receive from those who observe us. Our audiences online reflect back to us what their opinion of our behavior is, and we adapt to fit it. The metrics (likes, followers, shares, and comments) available to us now on social media allow for us to measure that feedback far more precisely than we previously could, leading to us internalizing what is 'good' behavior … Anytime we post to our followers, we are entering into a process of exchange with our viewers—one that is beholden to the same extreme engagement problems found everywhere else on social media." (2023, 145)

can erode authenticity in discourse by pressuring actors into appeasing their audiences and speaking and behaving in whichever manner happens to be most profitable at a given moment.

Let us now consider the social good of equity in accountability. This is the third and final social good that will be explored in this discussion. It was argued in the previous chapter that while people who possess significant economic privilege can be insulated from the chilling effects associated with social punishment, people who are less economically secure are much more likely to engage in self-censorship for the sake of ensuring that their material needs are met. The question we now face is whether social media in particular has the power to undermine equity in accountability via its propensity to fuel intimidation. My contention is that the dynamics of social media are much more likely to be costly to the intellectual and expressive freedom of people who possess fewer economic advantages than those who are more fortunate with respect to the resources at their disposal.

al-Gharbi has presented powerful arguments for the view that cancel culture, broadly construed, is a phenomenon that is more threatening to vulnerable members of society than it is to those who occupy positions of power and prestige:

> Defenders of what has come to be referred to as 'cancel culture' often attempt to portray the phenomenon as folks from less advantaged backgrounds holding the 'privileged' to account. In fact, the people engaged in these practices are typically themselves elites or aspiring elites. Again, symbolic capitalists tend to be among the most sensitive and most easily offended sectors of U.S. society. It is people like us who tend to be 'very online,' who focus intensely on race, gender, sexuality, and politics, and who take part in online mobbings. It is *elites* who are raised from a young age to understand and learn how administrative systems and processes work, allowing them to know which levers to pull to get people fired or disciplined, even on false or exaggerated charges, while minimizing repercussions or blowback for themselves. It is *elites* who feel comfortable folding authorities and third parties into their personal disputes, believing that these institutions, processes, and professionals exist to serve *their* interests (not wrongly), and that the system will typically

> work to their advantage (not wrongly). It is people from elite backgrounds who simply expect institutions and their representatives to accommodate their personal preferences, priorities, and perspectives … These kinds of knowledge, dispositions, and behaviors toward institutions are part of the 'hidden curriculum' of elite childhoods, elite education, and elite culture. Consequently, while there are many cases of elites 'canceling' working-class people, there are *not* many cases of nonelites successfully canceling elites. Even in the cases of 'punching up,' what is characterized as 'holding the privileged to account' is generally an instance in which some faction of elites has managed to purge or inflict damage on someone even better positioned than themselves. Much like cricket or lacrosse in the United States, cancellation is primarily an elite sport. (2024, 278)

al-Gharbi's account is sensible and compelling, and I wish to supplement it by pointing out some ways in which the dynamics of social media are more favourable towards members of society who occupy positions of privilege than members of other demographics that are less fortunate. It is plain to see how the dynamics of social media platforms can have differential impacts on these populations. For example, it is uncontroversial to point out that people who are in possession of significant affluence generally have more power to influence public opinion than their less affluent counterparts. If necessary, people in positions of privilege can mobilize the resources at their disposal in order to shield themselves from online controversies and avoid the most painful aspects of social punishment. In extreme cases, people can hire public relations firms and social media managers to help them repair their image after becoming a target of online castigation. If needed, these people can invest in high-quality photo and video content that will capture the attention of social media audiences and convey a message that is beneficial to them. While these tactics may not always produce their desired outcome, the point still stands that privileged individuals and groups can protect themselves from online attacks in ways that less fortunate people are unable to. It is often argued that it is unfair for affluent people to deploy their resources in order to receive much better legal representation in court proceedings than those offered to members of the public who are impoverished. While I will not take on this issue here, it is fair to point out that similar

concerns about inequity can plausibly be raised in discussions about online intimidation culture and its impact on different communities.

While access to financial resources is obviously an important factor in discussions about online intimidation culture and equity in accountability, it must be noted that fame and social status can also confer benefits on people who are targeted with attacks on social media. In some cases, people in high-profile positions who have amassed significant followings throughout their lives can rely on their supporters to come to their defence if and when they become subjects of online controversy. While fame can make one an easy target for social media backlash, it can also function in one's favour as social media audiences who feel invested in a particular high-profile person choose to spend time defending this individual from the castigation and accusations that abound on social media. Meanwhile, people who lead private lives and have no significant public profile may struggle greatly to shield themselves from online attacks if and when they find themselves in the proverbial crosshairs of social media mobs.[92] While high-profile individuals can use the power of modern media to inoculate themselves from some of the more damaging aspects of online intimidation culture, low-profile individuals are unlikely to successfully ward off online attacks unless powerful people and institutions who are sympathetic towards them choose to provide them with a platform through which they can spread their message.

Much more could be said about the relationship between online intimidation culture and equity in accountability. We can point out that economically advantaged people have greater

[92] Nathan P. Kalmoe and Lilliana Mason offer relevant commentary: "… political aggression appears to be aimed more at ordinary people, even when famous targets are accessible on social media and via public information … when we ask about abusive behavior, we see that people are more interested in targeting their fellow citizens." (2022, 97)

access to social media in the first place, and tend to develop stronger presences on social media platforms than others who are less fortunate. Indeed, al-Gharbi points out that "[f]requent social media users tend to look a lot like heavy podcast streamers: young, highly educated, and relatively affluent." [93] It seems that much of the time, those who are less privileged must direct their time and energy towards generating enough income to meet their basic needs, and do not always have the time and energy required to familiarize themselves with the complex and ever-changing social media landscape. In addition, we can point out that social media is a highly visual medium that is much more congenial to people who are conventionally physically attractive than those who are less advantaged in this area.[94] We can point out that many of the controversies that take place on social media involve disputes over language and symbolism that are the purview of the highly educated, and are relatively alien to many working-class people.[95] The overarching point is that the dynamics of social media clearly have the power to undermine the social good of equity in accountability, just as they have the ability to undermine critical intellectual faculties and authenticity in discourse. If we believe that these social goods are worth preserving, then it follows

[93] The passage continues: "For virtually all social networks, those with college degrees, with incomes over $75,000, or who live in urban areas are the most likely to use social media—and they tend to engage with these platforms much more frequently than other users …" (2024, 193)

[94] John D. Boy and Justus Uitermark write: "Although the specter of 'virtue signaling' looms particularly large when political beliefs are at stake … it indicates a broader dilemma. Instagram users are haunted by the question of whether their posts reflect who they really are or what others want to see. Our interviewees paid very close attention to how their posts were received, monitoring how many comments and likes they generated. Faces and bodies, especially beautiful faces and bodies, do well." (2023, 59)

[95] al-Gharbi argues the strict rules surrounding language and symbols are especially costly to the less privileged: "…today many symbolic capitalists seem to attribute *too much* power to symbols, rhetoric, and representation … Under the auspices of preventing these harms, they argue it is legitimate, even necessary, to aggressively police other people's words, tone, body language, and so forth. As we have seen, people from non-traditional and underrepresented backgrounds are among the most likely to find themselves silenced and sanctioned in these campaigns…" (2024, 297) Messina makes a similar point: "…elites often effectively punish uneducated people who lack the tools for understanding why their behavior is irresponsible. This can lead to conversations being dominated by the well-educated … one of the more sensible complaints about the often shifting goalposts of political correctness is that it allows elites to decide that an ever-narrower range of expressive acts is acceptable. Those not sufficiently initiated into elite circles don't have a real chance to participate. This is an issue because discourse is more productive when more people of diverse educational and socio-economic backgrounds contribute to it." (2023, 57)

that it is appropriate to seek changes to our media ecosystem so that online intimidation culture can be reined in, and its corrosive effects can be mitigated over time.

iii.vi: The Resilience of Societies

The previous chapter offered a brief definition of a "social good" for the sake of clarifying what this language entails in the context of this discussion of social media and free expression. It was noted that in order to qualify as a social good, a feature of social life must reliably advance utility as well as the long-term resilience of societies. At this point, I wish to say more about the issue of resilience, and why it is plausible to think that a Millian atmosphere of free expression has the potential to assist societies in thriving over the long term. My contention is that while talk of free expression and its associated social goods may sound lofty and distant from the everyday concerns that ordinary people face, there are good reasons to worry about online intimidation culture making societies less resilient in the face of serious challenges. When public discourse atrophies, so too does society's ability to develop a thorough understanding of problems, and to develop responses to these problems that can minimize human suffering and maximize human flourishing.

In order to better appreciate why it is plausible to think that healthy public discourse can assist societies in remaining resilient over the long term, let us consider the words of legal scholar Cass Sunstein. Sunstein argues that self-censorship can generate enormous losses at the societal level, and that resisting the pressure to engage in self-censorship can be a praiseworthy act:

> Under certain conditions … self-censorship is an extremely serious social loss. For example, Communism was long able to sustain itself in Eastern Europe not only because of force but also because people believed, wrongly, that most people supported the existing regime. The fall of Communism was made possible only by the disclosure of privately held

> views, which turned pluralistic ignorance into something closer to pluralistic knowledge … self-censoring can undermine success during war. Reputational pressures also help fuel ethnic identifications, sometimes producing high levels of hostility among groups for which, merely a generation before, such identifications were unimportant and hostility was barely imaginable. And if certain views are punished, unpopular views might eventually be lost to public debate, so that what was once 'unthinkable' is now 'unthought.' Views that were originally taboo, and offered rarely or not at all, become excised entirely, simply because they have not been heard. Here too those who do not care about their reputation, and who say what they really think, perform a valuable public service, often at their own expense. (2021a, 70-71)

Sunstein goes on to explain how a system of free expression can be extremely beneficial even to members of society who have little interest in directly taking advantage of the expressive freedom that is afforded to them. He states:

> Various civil liberties, including freedom of speech, can be seen as an effort to insulate people from the pressure to conform, and the reason is not only to protect private rights but also to protect the public against the risk of self-silencing … a system of free speech confers countless benefits on people who do not much care about exercising that right. Consider the fact that in the history of the world, no society with democratic elections and free speech has ever experienced a famine—a demonstration of the extent to which political liberty protects people who do not exercise it. (2021a, 71)

These comments suggest that an atmosphere of free expression can prove empowering for societies as they identify and respond to internal and external threats. Serious challenges such as ethnic strife, military aggression, and economic precarity are more likely to be addressed successfully when broad swaths of the population are given political and social permission to express themselves in a genuine manner and enter into meaningful dialogue with others. Sunstein's commentary provides support for the Millian conviction that the human intellect is an enormously powerful engine of social progress, and that the stifling of intellectual and expressive freedom, which are intimately linked, can inflict damage on society that extends far beyond any specific individual or group that happens to be a direct target of censorship at a given moment. If the arguments presented by Mill and Sunstein are sound, then it follows that humans' capacity for

reason ought to be conceptualized as an exceptionally powerful asset that can play a role in comprehending and solving a vast array of serious problems that have the potential to undermine the stability of society.

At this juncture, it is appropriate to note that each society possesses a certain amount of intellectual capital. While societies do have the power to bolster their supply of intellectual capital by constructing institutions that are conducive to learning and good-faith debate, it is nonetheless the case that no society possesses an infinite amount of energy and talent. These are perishable resources, and therefore, they ought to be allocated in a manner that is conducive to the wellbeing of society and the individuals that comprise it. It should be uncontroversial to suggest that while there are many worthwhile ways in which intellectual capital may be deployed, paramount among these is the project of addressing serious problems that afflict society, or that can be expected to afflict society in the future. If we accept the premise, following Mill, that human ingenuity has enormous power to identify solutions to problems and make life better for members of our species, then it follows that this powerful resource ought to be allocated in a responsible manner. Some projects and objectives are simply more important and deserving of resources than others. It is an unfortunate reality that in many cases, intellectual capital is squandered as the finite time, energy, and talent of human beings are directed towards goals that are either of no benefit to society or are actively harmful towards it.

These comments, which I view as being fairly straightforward and commonsensical, bring us to one of the most disturbing features of the modern phenomenon of online intimidation culture: namely, that this phenomenon entails tremendous opportunity costs. The pressures of online

intimidation culture motivate many people to invest exorbitant amounts of time, energy, and talent towards the goal of securing their own social standing and avoiding the pain of censure and exclusion. It is impossible to know with certainty how many hours humans have collectively spent playing the proverbial game of social media for the sake of reaping social rewards and evading social punishment, but we can know for sure that the number is far from negligible. Simultaneously, the pressures of online intimidation culture have clearly motivated others to retreat from public discourse for the sake of distancing themselves from the toxicity that pervades social media discussion, and can easily spill over into offline settings. In the former case, intellectual capital is deployed in a manner that is wasteful, and in the latter case, intellectual capital is simply neglected and permitted to wither away as people with considerable promise are marginalized and prevented from participating in intellectual exchanges that could be beneficial for themselves and for others. When we seriously contemplate how online intimidation culture has acted as a drain on society's finite supply of time, energy, and talent, it becomes clear that enormous gains can by generated by reforming our information environment so that the intellectual capital that is available to our society can be directed towards more productive ends.

While Mill does not deploy the language of "intellectual capital" in his writings, he does evoke a similar idea in the final passage of *On Liberty*:

> The worth of a State, in the long run, is the worth of the individuals composing it; and a State which postpones the interests of *their* mental expansion and elevation, to a little more of administrative skill, or of that semblance of it which practice gives, in the details of business; a State which dwarfs its men, in order that they may be more docile instruments in its hands even for beneficial purposes, will find that with small men no great thing can really be accomplished; and that the perfection of machinery to which it has sacrificed everything, will in the end avail it nothing, for want of the vital power which, in order that the machine might work more smoothly, it has preferred to banish. (2015, 106)

This passage functions as a warning about the potential for conformist pressures to deprive societies of the ability to thrive over the long term. Mill is fearful of living in a culture wherein people are encouraged to be "docile" and "small", viewing this as corrosive to society as a whole. It is understandable why Mill would articulate this concern, given his views about the enormous power of the human intellect to generate innovation and social progress over the long term.

I wish to expand on Mill's warning by arguing that while it is true that a culture of conformity can undermine intellectual capital by causing less of it to be present within a society, it is also true that societies that possess plenty of intellectual capital can deploy these resources in petty and unproductive ways when perverse incentives are present in public discourse. Mill expresses concern about "small" people interfering with societal flourishing, and accordingly, it makes sense to point out that there is more than one way for people to be "small". While people can be small in the sense of being intellectually impotent and lacking the wherewithal to grapple with difficult issues in a skilled manner, they can also be small in the sense of directing their time, energy, and talent towards projects that are shallow and frivolous, providing no genuine benefit to themselves or society more broadly. While it is reasonable to worry about societies becoming intellectually stagnant, it is also reasonable to worry about societies incentivizing their members to channel their finite resources towards goals and battles that are ultimately unproductive. One reason why it is appropriate to view online intimidation culture as a pressing issue, rather than merely an inconvenience or an irritant, is that this phenomenon directs people's time, energy, and talent away from important matters and towards trivial matters.

We should not lose sight of the impact that this has on society in the aggregate. Online intimidation culture is interfering with the ability of societies to respond appropriately to various threats that currently exist, as well as threats that lie on the horizon.[96] In a world that is increasingly complex, interconnected, and prone to rapid change, it is arguably more important than ever for public discourse to function in a manner that enables large populations to understand extant and anticipated threats to society and the myriad ways in which these threats may be addressed.[97] Public discourse is a crucial mechanism for sorting error from truth and identifying solutions to problems, and it cannot function properly when vast swaths of the population are preoccupied with the never-ending project of impressing others and obtaining social status, or alternatively, opt out of participation in public discourse because they are fearful of actors that are more aggressive and extreme in character. This mechanism is undermined when conversations about serious issues that impact vast populations are crowded out by skirmishes that, while potentially exciting to witness, do nothing to edify human beings or improve their station in life in any tangible way. This means that an atmosphere of free expression and its associated social goods should not be viewed as luxuries that are divorced from the everyday concerns of laypeople. Rather, they ought to be viewed as forces that can play a critical role in enabling societies to unlock the power of public discourse so that they can remain resilient in the face of various serious challenges and enjoy gains over the long term. The chapters that follow will explore a variety of strategies for addressing the

[96] Walter argues that the influence of social media has the power to undermine democratic societies: "People don't realize how vulnerable Western democracies are to violent conflict. They have grown accustomed to their longevity, their resilience, and their stability in the face of crises. But that was before social media created an avenue by which enemies of democracy can easily infiltrate society and destabilize it from within." (2022, 124)

[97] Interestingly, Macedo and Lee invoke Mill's philosophy in their discussion of the censorious climate that arose during the COVID-19 global pandemic: "If a national conversation about the pandemic is ever going to take place, now is the time for that conversation, so we can confront the fractures that the Covid crisis revealed in our basic democratic and scientific norms … In the twenty-first century, would one have believed that stigmatization of dissent—precisely as described by John Stuart Mill in 1859, before the U.S. Civil War—would still be a recurrent feature of our liberal democratic institutions? That government officials would engage in an active effort to censor their political opponents for expressing dissenting views?" (2025, 297)

problem of online intimidation culture, and offer an assessment of the extent to which these strategies cohere with the normative Millian vision that informs this discussion.

Chapter iv: Strategies for Addressing Intimidation Culture

iv.i: The Need for a Response

The preceding chapter has advanced the argument that online intimidation culture is a genuine phenomenon that undermines free expression, as well as the social goods that a system of free expression helps generate. It has been argued that while online intimidation culture is not a form of formal, political censorship wherein actors are punished by states for falling afoul of rules surrounding speech, it does amount to a form of informal, social censorship wherein various actors are stifled and pressured into conformity via fear of social punishment. The question that animates this chapter is how those who are concerned about this phenomenon might think about responding to it. It is one thing to draw attention to a pernicious trend in society, and it is another to identify strategies that might prove useful in efforts to combat it. This chapter will aim to give a fair hearing to a handful of different ideas about how those who care about free expression might go about addressing the tendency of social media to generate outrage, personal attacks, and the chilling effects that accompany them. I will begin by considering the proposition that online intimidation culture can best be addressed through a government ban of social media platforms. Then, I will explore the topic of government regulation, and offer commentary about whether regulation of social media companies should be viewed as the appropriate antidote in this area. In addition to considering strategies that involve the deployment of state power, I will assess the notion that the best way to address online intimidation culture is for consumers to voluntarily exit the realm of social media rather than for governments to take the lead in this area. I will conclude by noting that while these approaches to addressing online intimidation culture do have varying degrees of merit, none of them amount to a panacea that has the power to eliminate the corrosive dynamics that pervade our contemporary information environment.

Importantly, it must be stated that the discussion of strategies for addressing online intimidation culture that is offered in this chapter, as well as the discussions in later chapters, proceed from a Millian, liberal perspective. Just as Chapters 2 and 3 invoked the philosophy of Mill in order to make the case that the patterns of communication that pervade social media threaten free expression and key social goods, this chapter will invoke Millian ideas in order to reach conclusions about whether a proposed strategy for addressing intimidation culture is appropriate. We are not looking for just any strategy for addressing the problem of online intimidation culture; we are looking for strategies that are capable of bolstering free expression and the social goods that free expression is responsible for producing. A strategy that successfully addresses the problem of online intimidation culture, but imposes other costs on society that are equally pernicious or more pernicious from a Millian perspective, will not be satisfactory for our purposes. If there is no perfect strategy for addressing the problem of intimidation culture, then it is incumbent upon us to determine which of the available strategies is most aligned with the liberal ideals and precepts that have been identified throughout the preceding chapters.

iv.ii: Government Bans of Social Media

One potential remedy to the problem of online intimidation culture, and one that is particularly radical, is for states to ban social media altogether.[98] It might be reasoned that since

[98] While it is practically unheard of for high-profile individuals to advocate for a complete ban of social media, TikTok in particular has become a target of proposed bans due its ties to the Chinese government. 2024 Republican Presidential Candidate Nikki Haley publicly supported a ban of TikTok (Vigdor and Cameron 2023), as has NYU business professor Scott Galloway (Galloway 2023). In 2024, President Joe Biden signed a law requiring TikTok to be sold by its parent company, ByteDance, or face a ban in the US. Commentators such as Geoffrey Cain of the National Security Institute of George Mason University spoke out in support of this action (Grafstein 2024). These events indicate that social media bans are more than a merely theoretical possibility. There is real, energetic interest in the project of banning social media platforms that are deemed to be hazardous to Western liberal democracies. Marietje Schaake notes: "…not long after TikTok bans were suggested in the United States, European authorities soon followed. The TikTok case is both exceptional and exemplary of the pitfalls of American tech regulation. American policymakers are hyperfocused on the national security segments of tech regulation while remaining downright apathetic on

the culture of social media undermines important social goods, it is appropriate for states to intervene in a bold manner so that these social goods can be protected. To be sure, if social media discourse is dissolved entirely, then so too will its ability to fuel intimidation. If a social media ban were successfully implemented, then we could expect the pernicious impact of these platforms to fade away as people are forced to seek out alternative venues for communication.[99] Indeed, some may argue for a ban on social media on the grounds that it is sometimes necessary for states to remove toxic products from the marketplace in cases wherein the market's regulation of itself is unsatisfactory. Many jurisdictions have banned the use of lead in paint and gasoline on the grounds that this substance is severely toxic, and inhibits the proper physical and mental development of humans.[100] Some may argue that just as lead is toxic in a literal sense, social media is toxic in a figurative sense, as it inflames social tensions and undermines toleration and cohesion.[101] According to this line of reasoning, if it is legitimate and desirable for states to ban lead in various contexts for the sake of the health of the populace, then it should also be seen as legitimate and desirable for states to ban social media in the interest of the greater good.

---

questions of civil liberties like data privacy. When national security appears to be at risk—as is the case around TikTok—U.S. politicians take dramatic action, often swiftly. Yet when tech overreach infringes on the rights of average Americans, lawmakers may write an op-ed or pen a press release, but they don't manage to take meaningful action through Congress." (2024, 187)

[99] Neta Kligler-Vilenchik and Ioana Literat specifically note: "With youth driving its meteoric rise in the past few years … TikTok … has also become significantly more political—and, in large part due to the ongoing attempts to ban it, more *politicized*—since the 2016 election. Indeed, the platform's central role in young people's political lives is now widely recognized and has also been the focus of an increasingly rich and diverse body of research." (2024, 65)

[100] Haidt mentions leaded gasoline in his critique of social media companies: "We can … compare them to the oil companies that fought against the banning of leaded gasoline. In the mid-20th century, evidence began to mount that the hundreds of thousands of tons of lead put into the atmosphere each year, just by drivers in the United States, were interfering with the brain development of tens of millions of children, impairing their cognitive development and increasing rates of antisocial behavior. Even still, the oil companies continued to produce, market, and sell it. (2024, 4)

[101] The technologist Jaron Lanier also provides an analogy between social media and leaded products: "When it became undeniable that lead was harmful, no one declared that houses should never be painted again. Instead, after pressure and legislation, lead-free paints became the new standard. Smart people simply waited to buy paint until there was a safe version on sale. Similarly, smart people should delete their accounts until nontoxic varieties are available." (2018, 29)

This is an interesting idea. There is no doubt that in some cases, banning a harmful product from the marketplace can indeed generate meaningful gains for ordinary people and bring about a healthier status quo. However, in addition to being aggressive, such a strategy for combating online intimidation culture would be overwhelmingly illiberal, and would raise concerns about free expression that dwarf the concerns enumerated in the previous chapter. If states were given a proverbial green light when it comes to using their power to purge social media from the marketplace, it is difficult to see why the project of eliminating "toxic" media should stop here. There are innumerable forms of media that one can plausibly argue have a pernicious impact on the quality of public discourse. If states were given license to shut down media platforms for the sake of reining in intimidation and shaping public discourse, then they could use this power in ways that people with a liberal orientation ought to find disturbing.[102] We could reasonably expect such states to target a broad array of media with censorship in order to silence those that challenge their interests and agenda, which would obviously be an affront to free expression as well as other basic liberties.

Moreover, a ban on social media would raise important questions about democratic legitimacy. While it is true that public awareness of the corrosive effects of social media has increased significantly in recent years, it is not the case that there is currently a major, grassroots effort on behalf of citizens to have social media banned in a wholesale manner. Many people view

[102] Zac Gershberg and Sean Illing offer an account of authoritarian policies surrounding media that were enacted in the Soviet Union: "Media had been tightly controlled throughout the Soviet era, not just in the closed production processes of state-sanctioned propaganda but also in terms of access to technology itself … What they feared, above all, was the proliferation of self-published samizdat." (2022, 183)

social media as a net positive in their own lives,[103] and many choose to continue using social media services while remaining vigilant with respect to shielding themselves from the more unhealthy dynamics that permeate these platforms. Simply put, there are many people who like social media platforms and want to continue using them. This is one area wherein the analogy between social media and leaded paint and gasoline begins to break down. When leaded products were banned by governments in the 1970s, there was no contingent of consumers that fought such bans because it believed that it stood to gain from these products. There was broad agreement with respect to the idea that these products were dangerous and needed to be removed from the marketplace. Such agreement is absent in the modern context involving social media. A government ban on social media would fly in the face of the desires of a large share of the population, meaning that it would be an elite-driven policy that lacks the approval and consent of the governed. Anyone who cares about democratic legitimacy must address this point if they wish to construct a plausible argument in favour of banning social media via the state.

Another point that undermines the notion that it is appropriate for governments to ban social media wholesale concerns the issue of agency in the marketplace. Let us once again examine the analogy between social media products and leaded products. When people are exposed to toxic substances such as lead, they have little, if any, ability to determine the way in which this exposure will affect their own physical and mental health. When lead enters the body, it causes damage, and reversing this damage can be extremely difficult. However, in the domain of social media,

[103] Vaidhyanathan notes how social media can enhance the lives of individuals while it corrodes society: "Facebook likely has been—on balance—good for individuals. But Facebook has been—on balance—bad for all of us collectively. If you use Facebook regularly, it almost certainly has enhanced your life. It has helped you keep up with friends and family members with regularity and at great distance. It has hosted groups that appeal to your hobbies, your interests, your vocations, and your inclinations. Perhaps you have discovered and enjoyed otherwise obscure books and music through a post from a trusted friend." (2021, 20)

consumers actually do play a role in determining how they will be impacted by exposure to toxicity, and there is little reason to think that exposure must be damaging in all cases. When confronted with social media content that is decidedly toxic in nature, users can choose to ignore it, to find amusement in it, or perhaps even learn from it. Social media users do not need to be made worse off every single time that they encounter social media content that is toxic, and importantly, their relationship with such social media content can evolve over time. An individual who is severely hurt and intimidated by toxic social media content in their teenage years may develop immunity to this content later in life, and may even develop a sense of humour with respect to it as a coping mechanism. The key point is that even if we accept the premise, as I do, that social media content can be seriously pernicious, it does not follow from this that banning social media is analogous to banning leaded products, as consumers have agency in the former context that is clearly absent in the latter context.[104]

Let us recall that one of the social goods that the preceding chapter has brought to the fore is that of critical intellectual faculties. A key reason that I have argued that intimidation culture is pernicious is its tendency to erode this important social good. It is worth asking whether banning social media is perhaps even more threatening to this social good than the climate of hostility and outrage that currently pervades online discourse. Even if we agree that hostile and outrageous content is damaging critical intellectual faculties by derailing public discourse that could otherwise

[104] It is worth noting that social media companies have age restrictions in place that officially prohibit persons below a certain age from becoming participants on the platforms that they provide. While reasonable people can disagree about whether such policies ought to be more or less strict than they are, the fact that age restrictions are widely accepted as legitimate indicates that there is broad agreement around the idea that users exercise some degree of agency in the realm of social media. We generally have a good intuitive understanding that people who are sufficiently mature and developed can expose themselves to a broad array of content without necessarily being adversely impacted by it, while people who are still in early stages of development are at greater risk. This reasoning applies in the domain of social media as it applies in others involving media.

be productive, it is not clear why invoking the power of the state to eliminate such content would be beneficial with respect to these faculties. Critical intellectual faculties are strengthened not by evading challenges, but by confronting them and striving to identify effective solutions. If the cacophony of social media is addressed simply by calling upon the state to shut down platforms, then this arguably amounts to an abdication of responsibility on behalf of people who can and should deploy their critical intellectual faculties to challenge the status quo in the realm of social media and work to cultivate a healthier information environment. A government ban on social media may incentivize and normalize complacency and idleness rather than the development of critical intellectual faculties, which is far from desirable.

While it is interesting to think about how a social media ban might shape society, this strategy for addressing the problem of online intimidation culture is not sensible or realistic. Although banning social media would, practically by definition, help eliminate the problem that motivates this discussion, it would raise other concerns that are even more significant from a Millian perspective. It would be foolish, and even ironic, to address the chilling effects of social media by implementing a policy that will surely produce chilling effects of its own that are far greater in scope and severity. This approach would effectively exchange social censorship for political censorship, thereby jeopardizing free expression and the social goods that it produces, instead of achieving progress in this area. If we are concerned about the chilling effects produced by social punishment in the realm of social media, as we ought to be, then it makes good sense to also be concerned about chilling effects produced by governments tasked with shutting down media outfits on the grounds that they are injurious to public discourse. In such a state of affairs, media outfits would be forced to self-censor in order to avoid formal state punishment, which

would of course be a net loss for free expression and the societal marketplace of ideas. Arguably, discussions about banning social media are primarily useful because they can bring liberal principles and social goods that are worthy of protection into clearer view. If we are interested in identifying strategies for improving social media that are consistent with such a perspective, then we must explore approaches that are more measured in nature. The remainder of this chapter will explore strategies for improving social media, and for addressing the problem of online intimidation culture, that are less heavy-handed, and more likely to garner popular support in the short term.

iv.iii: Government Regulation of Social Media

A more nuanced approach to the problems raised by social media may involve government regulation of social media platforms rather than an outright ban. While those with a liberal perspective must remain cognizant of the risks involved in ceding power to the state with respect to regulation of the media marketplace, there is little reason to think that regulation will invariably function in a manner that stifles free expression. Perhaps regulation that is intelligently crafted and properly targeted can actually help create and sustain an atmosphere of open debate and inquiry instead of undermining it. Indeed, some existing legal constraints on expressive acts arguably function in this manner. By their nature, defamation laws place limits on what can be legitimately said and published about people. However, it is generally accepted that such laws have a positive impact on individual liberty and the quality of public discourse because they offer people an avenue for recourse if and when they are targeted with malicious falsehoods.[105] Accordingly, such

[105] Sunstein explains that protecting one's reputation can be intertwined with their liberty: "One of my concerns is people's ability to protect their reputations. Your reputation can be seen as part of your property and as one of your liberties. It is no light thing to take someone's property or to diminish their liberty." (2021b, 8) He later explains that

laws disincentivize the circulation of malicious falsehoods as actors come to understand that if they engage in such behaviour, serious repercussions can follow. Rather than stifling discourse, laws against defamation can help liberate it by giving many different individuals and groups an opportunity to express themselves without living in fear of having their reputations and livelihoods ruined. It is noteworthy that while there are currently many disagreements between actors across the ideological spectrum regarding how we ought to think about free expression, there is generally a stable consensus around the idea that prohibitions on defamation are legitimate, suggesting that it is possible to reach agreement in this area even in times of pervasive polarization.[106]

Similar logic may apply in the realm of social media: perhaps instead of undermining free expression, well-crafted legislation can actually help unlock the potential for social media platforms to function as productive venues for diverse and heterodox thought.[107] Accordingly, Millian liberals ought to carefully engage with proposed regulations rather than dismissing them in a kneejerk fashion on the grounds that they are injurious to free expression. If a plausible case can be made that proposed regulation in the social media marketplace will help to facilitate meaningful communication between individuals and groups with many different worldviews, then it is inappropriate to cling to the conclusion that it flies in the face of liberal ideals. It is worthwhile to note that while Mill obviously does not offer arguments about experimentation in the realm of social media in his seminal texts, the value of experimentation is a theme that manifests itself throughout his work. Mill repeatedly points out that trial and error is a powerful mechanism for

laws against defamation dually serve to discourage defamation and to provide opportunities for relief after defamation has taken place. (2021b, 90)

[106] This is not to suggest that people with varying ideological orientations generally agree on cases involving charges of defamation. The point is that in principle, people generally accept that it is legitimate for states to place limits on expression on the grounds that it is defamatory.

[107] Taylor Owen and Supriya Dwivedi hold that "democratic platform regulation can maximize free speech in a way that the market is unable to do." (2022)

the discovery of truth and the identification of new strategies for improving human affairs.[108] Millian liberals[109] ought to remain open to experimentation in the private sector, as well as in the public policy arena, in the interest of improving social media and the discourse that it facilitates. Given the overwhelming ubiquity and influence of social media, it would be an error to shut ourselves off from promising ideas that arise in the private and public domains. Even if we accept the notion that government regulation can erode the dynamism of the social media marketplace and undermine free expression, it does not follow from this that all proposed regulation ought to be dismissed without analysis. This would simply be dogmatic and unwise. In some cases, regulation may even enhance the ability of firms in the social media marketplace to engage in productive experimentation, which ought to be welcomed.[110]

There are many ideas circulating in academia and popular media about how social media companies can be better regulated, and not all of them can be discussed in detail here. Some commentators argue that social media companies should bear fiduciary responsibilities with respect to their users,[111] similar to the manner in which parents and medical doctors are bestowed

[108] Consider the following passage from *On Liberty* wherein Mill connects the concept of experimentation with "individual and social progress": "As it is useful that while mankind are imperfect there should be different opinions, so it is that there should be different experiments of living; that free scope should be given to varieties of character, short of injury to others; and that the worth of different modes of life should be proved practically, when any one thinks fit to try them … in things which do not primarily concern others, individuality should assert itself. Where, not the person's own character, but the traditions or customs of other people are the rule of conduct, there is wanting one of the principal ingredients of human happiness, and quite the chief ingredient of individual and social progress." (55-56)

[109] Messina notes that there is a strong connection for Mill between free expression and the process of experimentation: "…in addition to its relationship to the truth, freedom of thought and expression are important for allowing us to envision and enact experiments in living, by which we depart from the common ways of doing things and carve out our own paths. The ways in which censorship can impede the development of these experiments is not merely bad news for our autonomous self-development and capacity to develop as individuals, it can also stop us from discovering problems in our local culture and better ways of doing things." (2023, 12).

[110] Bans on non-compete clauses in the social media sector are an example of the type of legislation that I have in mind here.

[111] Jack M. Balkin explains: "My own contribution to these issues is the concept of information fiduciaries. I've argued that the digital age has created great asymmetries of power and knowledge between the digital businesses that collect data from end users and the end users themselves. These asymmetries of power and knowledge create special

with such responsibilities. Others argue that stricter rules surrounding data collection can help rein in the influence of targeted advertising in the realm of social media, which motivates platforms to deliver a user experience that is maximally addictive.[112] The European Union's Digital Services Act requires, among other things, that social media companies share information with regulators about how their algorithms operate.[113] This can help demystify the issue of social media content moderation and help people outside of the technology industry understand which types of content are being emphasized and deemphasized in the world of online discourse. All of these techniques for regulating social media are logical to a certain extent, but it remains true that we are still quite distant from having conclusive evidence that they are beneficial for our purposes: whether or not these regulatory techniques can function as practical, rather than theoretical, remedies towards the pervasive problem of online intimidation culture remains very much an open question.

There are two specific techniques for regulating social media via the state that I wish to consider in detail in this chapter. The first concerns the issue of content amplification on social media, and the second concerns the issue of revenue generation. While there is no shortage of ideas about how social media platforms can be improved via government regulation, I have chosen to

vulnerabilities for end users that are the traditional concern of fiduciary law. Therefore, I've argued that businesses that collect data from end users must assume fiduciary duties of confidentiality, care, and loyalty to the people whose data they collect and use." (2022, 249-250, Bollinger and Stone, ed.) Tristan Harris, co-founder of the Center for Humane Technology, offers the following during a 2019 podcast: "You have to have a responsibility to the community that you are inside of and serving … so that's why we just need to just bite the bullet here and switch to a fiduciary model, and that's the biggest, most powerful action that government can help make possible." (Harris 2019)

[112] Taylor Owen and Supriya Dwivedi state: "…if our data were better protected from unfettered third-party use, then we could be spared from the foibles of the targeted advertising market, including the microtargeting of content that can be used to enrage and divide us." (2022)

[113] An official EU website states: "Algorithmic systems affect our experiences online. The Digital Services Act (DSA) is a legislative initiative by the European Union aimed at making the internet safer and protecting people's rights online. Algorithmic transparency and accountability are key parts of this protection. The European Centre for Algorithmic Transparency (ECAT) contributes with scientific and technical expertise to the European Commission's exclusive supervisory and enforcement role of the systemic obligations on designated Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) provided for under the DSA."

analyze these two in particular for several reasons. First, these techniques are structural in nature. Rather than focusing on a concerning type of content that pervades social media and directly calling for a stop to it, these approaches focus on the underlying incentives that are at play in the context of social media discourse.[114] If we are genuinely interested in the project of improving social media communications and facilitating a shift away from online intimation culture, then it makes sense for regulatory interventions to focus on incentives rather than specific categories of expression. The latter course of action is analogous to treating symptoms that manifest themselves when illness is present, while the former course of action is analogous to treating the root cause of an illness. If regulators can successfully realign the incentives that are at play in social media interactions, then they can reasonably expect to generate significant changes to the tone and content of online discourse, thereby challenging intimidation culture and making way for new patterns of expression and interaction that are more prosocial and constructive.

Another reason that regulatory techniques focused on content amplification and revenue generation are being brought to the fore in this discussion is that they are (arguably) more specified and determinate than other alternatives that are available in this area. For example, while I am not necessarily an opponent of the notion that social media companies should be required to disclose

[114] This is arguably one of the flaws that afflicts the Canadian federal government's Online Harms Act, which was unveiled in 2024 and attracted criticism from members of parliament and other commentators. This is a sweeping piece of legislation that deals with an array of contentious issues such the non-consensual spread of intimate images, terrorism, hate speech, and more. Indeed, the legislation grapples with a variety of forms of expression that fall afoul of Mill's harm principle, and are accordingly legitimate targets of government censorship from a Millian perspective. Mill's liberal framework does license governments to prosecute and punish certain forms of expression, specifically on the grounds that they are responsible for inflicting harm. The issue here is that despite its immense scope, this legislation offers very little in the way of structural changes to social media. The legislation does not explore the issue of incentives in a meaningful way. Instead of confronting the question of why social media platforms are generating so much pernicious content in the first place, the legislation seeks to implement tougher penalties for the circulation of such content. There is an important difference between regulatory approaches that attempt to reduce the creation and spread of damaging content by addressing the underlying incentives that shape it, and those that attempt to address damaging content by administering punishments to guilty parties after the fact. While the former approach is more proactive, the latter is more reactive.

information about their algorithms, or the notion that these companies should bear fiduciary responsibilities, it remains unclear how exactly these measures would generate tangible and desirable results with respect to the issue of online intimidation culture.[115] With this being the case, it seems appropriate to focus instead on techniques that have a clear and identifiable connection to the incentives that animate online discourse. Such techniques are more relevant to the pernicious patterns of communication and behaviour that were described in the previous chapter. The most important reason that I have chosen to focus on these two approaches is that they simply seem, upon consideration, more promising and impactful than the other approaches that have been proposed in academic and popular discussions. If we are going to take seriously the idea that online intimidation culture can be mitigated via government regulation of social media, then it is appropriate to focus our attention and energy on the techniques involving government regulation that appear the strongest. This does not preclude us from considering new methods that might arise in the future or novel ways of formulating methods that have already been recommended in various venues. Since it is not possible to examine all available methods here, it is reasonable to highlight ones that appear most viable with respect the specific issue of online intimidation culture, which motivates this entire discussion.

The first technique for addressing online intimidation culture that I wish to explore involves making social media companies legally liable for content that is circulated on their platforms if

[115] Bestowing social media companies with fiduciary responsibilities raises more questions than it answers. If a social media user spends a great deal of time consuming news content on a platform, and this content is damaging to their wellbeing due to the disturbing nature of the information that it includes, should this be viewed as a breach of fiduciary duty by the relevant social media company? Will individual employees at social media companies (such as content moderators) be bestowed with fiduciary duties, or will these duties apply to the organization as a whole? If a social media user is on the receiving end of ill-intentioned messages from other users, will the relevant company be required to block these messages in order to protect the user from pain and distress? These are some of the questions that would need to be addressed if a fiduciary model is embraced.

and when this content is amplified via content moderation practices. We may call this the "amplification regulation" technique for the sake of clarity. At present, thanks to pieces of legislation such as Section 230 of the Communications Decency Act in the United States, social media companies do not bear legal liability for the content that social media users post online.[116] If users take to social media to post content that is threatening, defamatory, or otherwise legally problematic, the offending party in such cases is the user and not the social media company in question. Amplification regulation would modify this arrangement so that social media companies lose their immunity when they boost content on their platform, thereby increasing its reach.[117] This would be the case regardless of whether the decision to boost the content is made by a human or a machine.[118]

The reasoning that underpins this technique is that amending the legal protections that social media companies receive would cause them to exercise much more caution with respect to content amplification. Indeed, evidence indicates that algorithmic curation has a major influence on the amount of traction that online content receives, meaning that revising algorithmic curation

[116] Jeff Kosseff explains the impact of this legislation on modern online communications: "Without Section 230, companies could be sued for their users' blog posts, social media ramblings, or homemade online videos. The mere prospect of such lawsuits would force websites and online service providers to reduce or entirely prohibit user-generated content … Consider the ten most popular websites in the United States as of 2018. Six—YouTube, Facebook, Reddit, Wikipedia, Twitter, and eBay— primarily rely on videos, social media posts, and other content provided by users. These companies simply could not exist without Section 230." (2019, 4)

[117] A report by Roddy Lindsay states that an opinion filed by Supreme Court Justice Clarence Thomas recommends that the US government should "preserve most Section 230 protections but eliminate them for algorithmically amplified content like that in Facebook's News Feed, which boosts the distribution of stimulating items that attract more clicks and comments." (2020) Kosseff, an advocate for Section 230 protections, concedes that some modifications to the law may be appropriate: " … we should all work to understand how to improve Section 230. Platforms must do a better job at blocking illegal or harmful third-party content, and if they are not doing that, then Congress should consider narrow carve-outs to Section 230 that address those problems without compromising the entire structure that the section supports." (2019, 280)

[118] Rose-Stockwell explains that algorithmic content curation plays a role in the proliferation of incendiary content: "For the first time, the majority of information we consume as a species is controlled by algorithms built to capture our emotional attention. As a result, we hear more angry voices shouting fearful opinions and we see more threats and frightening news simply because these are the stories most likely to engage us. This engagement is profitable for everyone involved: producers, journalists, creators, politicians, and of course, the platforms themselves." (2023, 33)

would result in significant changes to the user experience on relevant platforms.[119] At present, while social media companies may be disincentivized from promoting toxic content insofar as doing so can generate public backlash, they have little to worry about when it comes to legal repercussions. Amplification regulation introduces another disincentive into this area that is arguably more powerful than public backlash, as it can carry stiff fines along with it, as well as other penalties. If this form of regulation were implemented, then social media companies would have another good reason to avoid amplifying incendiary content that is rife with personal attacks. If these attacks are found to be defamatory, or legally problematic for another reason, social media firms will not be able to evade accountability simply by pointing out that the content was posted by users rather than the personnel of the relevant company. Their role in broadening the reach of the content will be the target of interest in such cases rather than their role in formulating the words or images that are presented to audiences.

This technique for alleviating the toxicity of online discourse is fascinating, and merits a good deal of discussion.[120] I am prepared to grant that amplification regulation can help make the social media ecosystem less hostile and divisive overall. Indeed, on an intuitive level, it seems wrong that gargantuan social media companies have been able to recommend incendiary content to users billions of times, helping it to capture the attention of users at the expense of other content that is more conducive to intellectual and emotional wellbeing, without facing any sort of serious

[119] For example, Hana Kiros notes "YouTube's recommendation algorithm drives 70% of what people watch on the platform." (Kiros 2022)

[120] Larry Kramer entertains the idea of limiting social media's ability to amplify content to audiences: "Why not limit the platforms' business models to prohibit pushing out certain kinds of information—not forbidding them to provide access, but simply disallowing the feature that lets them put unrequested information in front of unwitting users? … We could, likewise, permit platforms to show users potential content while following Roger McNamee's smart suggestion to ban algorithmic amplification." (2022, 38, Bollinger and Stone, eds.)

legal responsibility.[121] Introducing legal liability into this arrangement could introduce much-needed scrutiny and reflection into the process of online content moderation as social media companies are forced to make more careful decisions about which content to promote and which to refrain from promoting. Rather than simply inundating users with content that is most likely to capture their attention, companies would be forced to examine content to make sure that amplifying it will not place them on the wrong side of a legal controversy.

While amplification regulation deserves to be taken seriously, it would be an error to view this technique as a comprehensive remedy to the problem of online intimidation culture that was detailed in Chapter 1 of this work. It must be noted that countless media outlets publish content that is incendiary and divisive even when they know perfectly well that they are legally liable for their words and actions.[122] In many cases, speakers and publishers are happy to flirt with the limits of what is legally permissible in order to achieve attention, notoriety, and financial gain. Indeed, outrageous media has been a lucrative business since long before the rise of social media, and the presence of regulation has not stopped various actors from participating in this industry.[123] The notion that amplification regulation can eliminate the problem of online intimidation culture is too optimistic, even if this kind of regulation does introduce more caution and scrutiny into the process

[121] According to Guillaume Chaslot and Tristan Harris, YouTube recommended videos from notorious conspiracy theorist Alex Jones at least 15 billion times before he was deplatformed. (Harris, 2019)

[122] Journalist Matt Taibbi comments about the willingness of news outlets to circulate outrageous content in pursuit of financial gain: "In 2016 especially, news reporters began to consciously divide and radicalize audiences … As Trump rode to the White House, we rode to massive profits. The only losers were the American people, who were now more steeped in hate than ever." (2021, 5) Taibbi affirms that news outlets make a concerted effort to deliver content to consumers that is maximally addictive: "There is a terror of letting audiences think for themselves that we've never seen before … Keep clicking, keep delving deeper into the argument, make it more and more your identity … Click on, watch, read, tweet, argue, come back, click again, repeat, do it over and over, rubbing the nerve ends away just a little bit each time. With each engagement, you're signing over more and more of your intellectual autonomy." (2021, 205)

[123] Yochai Benkler notes: "Since the late 1980s, selling right-wing outrage has been big business, and its commercial success enabled it to take over the conservative media ecosystem." (2022, 255, in Bollinger and Stone, ed.)

of boosting online content on behalf of very large social media companies. Even if we embrace this proposal and agree that it is a step in the right direction, we must be prepared to consider additional methods for addressing the issue of online intimidation culture. At best, this technique is one small part of a more comprehensive solution.

The other regulatory technique that I wish to consider involves states requiring social media platforms to make major changes to their business model. We can refer to this as the "revenue regulation" technique. As many commentators have noted, much of the dysfunction on social media is attributable to the fact that social media companies generate revenue through advertising.[124] For this reason, they deliver a user experience that is highly addictive in order to keep users returning to the platform as frequently as possible.[125] There is an important connection between the toxicity of social media and the desire of social media companies to maximize their revenue from advertisers. Accordingly, states could order social media platforms to transition to an alternative model, such as a paid subscription revenue model, in order to realign the incentives in online spaces in hopes of improving the quality of public discourse.[126]

[124] Lanier explains: "… with old-fashioned advertising, you could measure whether a product did better after an ad was run, but now companies are measuring whether individuals changed their behaviors, and the feeds for each person are constantly tweaked to get individual behavior to change. Your specific behavior change has been turned into a product. It's a particularly 'engaging' product not just for users, but for customers/manipulators, because they worry that if they don't pay up, they'll be left out in the cold." (2018, 28)

[125] Sinan Aral states: "Engagement keeps our attention, which is what Facebook and all social media companies sell to advertisers. Newsfeed algorithms give us some diversity to explore the space of our preferences and keep things fresh and dynamic, but more than anything, they give us more of what we want, based on what we engaged with in the past. Ad-targeting algorithms maximize click-through rates, conversion rates, and customer lifetime value. (2021, 118)

[126] Rose-Stockwell discusses why the advertising revenue model has become dominant in digital media despite alternatives being available. (2023, 24)

I am sympathetic towards the idea that a subscription revenue model can be healthier than an advertising revenue model with respect to the incentives that it puts in place. When consumers are paying to access content, this decreases pressure on the relevant platform to deliver an experience that will keep consumers returning to it in a compulsive manner. This is because when consumers purchase subscriptions in order to access content, the providers of this content do not reap additional profits by enticing subscribers to remain glued to this content for as much time as possible. As long as users are willing to continue paying for their subscription over time, the content provider will continue to benefit financially. In contrast, the advertising revenue model that is today ubiquitous in the social media sector incentivizes firms to keep users returning to their platform as often as possible, and for as much time as possible. Rather than providing an experience wherein consumers have their attention captured to the maximum extent on a daily basis, a subscription revenue model encourages media companies to provide an experience that consumers will find valuable enough to warrant a financial commitment over the long term. A transition to subscriptions in the realm of social media could shape the content that users are exposed to in beneficial ways, as social media companies may have less incentive to amplify content that elicits a strong emotional reaction among audiences.

While this line of reasoning has some merit, the notion that a subscription revenue model will eliminate the pernicious patterns we now see on social media has not yet been vindicated empirically. X (formerly Twitter) now strongly encourages its users to purchase subscriptions in order to receive a verification badge on their accounts, and this platform remains as vitriolic and dysfunctional as ever. There is no sign that users who pay for this service behave in a less antagonistic manner than others who access the platform without a monetary payment. Moreover,

the contemporary media marketplace has no shortage of outlets that are supported through direct payments from their audience rather than advertisers. While these outlets are very diverse, and some of them do make a sincere effort to publish content that is thoughtful and measured rather than tribal and incendiary in character, others are overtly partisan[127] and are more than happy to provide their subscribers with outrageous content that will affirm their worldview. A great deal of profit can be made by producing content that makes audiences feel righteous, and encourages them to be judgmental and blameful towards others with different worldviews. While a transition from an advertising revenue model to a subscription revenue model may indeed entail some benefits with respect to the quality of public discourse, this technique cannot be counted on as a remedy that will put a stop to intimidation culture and enable people with many different kinds of ideas to interact with one another in a constructive manner.

Another risk associated with revenue regulation that ought to be recognized is that ordering a transition to a new business model will likely make it exceedingly difficult for new firms to enter the social media marketplace that can compete with ones that are currently dominant. If we want to see the social media marketplace improve over time and become less confrontational and divisive, then we ought to cultivate a marketplace that is hospitable to new entrants and does not punish small firms that are interested in experimentation and innovation. Messina offers insightful analysis about this issue:

> … regulation that applies to the entire industry will limit the ways in which future [social media platforms] can experiment with regimes of content moderation. In turn, this will limit consumers' abilities to choose the communities they wish to join. Additionally, the costs of complying with these sorts of regulations will not impact new entrants and existing players equally. Those that already enjoy networks of users and massive budgets can more readily absorb them than new entrants. Even if the regulatory solutions are initially

[127] Some examples of partisan online outlets that are supported through paid subscriptions are Fox Nation, DailyWire+, and BlazeTV+.

> narrowly tailored to exempt new entrants, entrenched interests can capture the regulatory bodies to the detriment of new entrants. But if [social media platforms] differentiate their products in part by providing different content curation and moderation services, we should not want to discourage new market entrants. (2023, 131)

Indeed, if new social media companies emerge in the marketplace that explicitly seek to offer consumers a more wholesome and prosocial alternative to the outrage and dysfunction that pervade existing platforms, and if consumers are sufficiently exhausted by the social media status quo that they are inclined to try out novel platforms that emerge in the marketplace, then these new entrants into the social media sector may effectively remedy the problem of online intimidation culture to a significant extent.[128] It would be a cruel irony if regulation were enacted in order to alleviate the hostility and division that dominate online discourse, only to exclude new companies from participating in the market that have the ability and willingness to help realize this objective. Those who are concerned about online intimidation culture ought to welcome the possibility of major social media platforms being supplanted by new, rival platforms that offer a user experience that is less conducive to toxicity and more conducive to good-faith interactions between users. Even if it turns out to be true that a transition away from advertising and towards subscriptions has positive implications with respect to social media discourse, this will be a pyrrhic victory if it means that new and potentially better social media platforms will be denied the opportunity to compete in the marketplace with their more established counterparts.

[128] DiResta points out that our social media ecosystem may be improved by new companies entering the marketplace: "…while it is easy to fall into the trap of assuming that the big-tech platforms that exist today will exist tomorrow, the emergence of new prosocial-first platforms, designed from the ground up, may be the way forward. There may be significant hurdles to the mass adoption of such platforms, but adjacent regulatory efforts … may create an opportunity for new entrants. (2022, 135)

It is also the case that governments ordering a transition away from the advertising revenue model will likely facilitate siloing of the social media ecosystem.[129] If consumers perceive a social media company as having an ideological bias, they may refuse to pay money for its service, and instead choose to support a rival company that is more congenial towards their own worldview.[130] Given how politicized the topic of social media content moderation has become in popular culture, and given the overtly partisan and antagonistic posture that X (formerly Twitter) in particular has assumed,[131] this must be viewed as a very real possibility rather than a merely theoretical one. It is already the case that former users of X (or Twitter) have chosen to migrate to alternative platforms such as Mastodon, Bluesky, and Threads in order to protest its leadership. If consumers with different political leanings continuously choose to use and support different social media platforms, then this will effectively eliminate opportunities for these groups to communicate meaningfully with one another in online spaces. Indeed, intimidation culture could be invigorated by siloing of the social media ecosystem as people with different political views are continually rewarded for proving their loyalty to their own ideological camp and launching attacks on opposing camps.[132] While it is true that social media companies could try to avoid this type of

[129] Gershberg and Illing describe how a divisive media environment fuels the sorting process: "We live in what Kuran calls two intolerant communities …These communities live in different worlds, desire different things, and share almost nothing in common. And these alternative universes are reinforced by a partisan media environment that delivers news like any other consumer product and sorts people into virtual factions … the solutions won't come merely from better legislation or institutional reforms or more virtuous politicians. We'll have to reestablish a healthy culture of democracy by improving the communication environment." (2022, 253)

[130] Stone expresses concern that echo chambers may become more prevalent as the information environment continues to change: "There's actually a reasonably well-established view now in political science that hardcore echo chambers are fairly uncommon … The literature's understanding of the prevalence of echo chambers may also change as media consumption data becomes more granular or due to the media landscape continuing to develop. For example, the growth of Substack subscriptions and partisan social media platforms … may cause bona fide echo chambers to become more common." (2023, 121)

[131] Twitter officially rebranded as "X" in 2023, but it remains common for people to refer to it with its former name.

[132] Sunstein argues that self-sorting can undermine freedom: "When people have multiple options and the liberty to select among them, they have freedom of choice, and that is exceedingly important. … But freedom requires far more than that. It requires certain background conditions, enabling people to expand their own horizons and to learn what is true. It entails not merely satisfaction of whatever preferences and values people happen to have but also circumstances that are conducive to the free formation of preferences and values … if people are sorting themselves

sorting by designing rules and content moderation policies that aim to appease people with various ideologies through fairness and impartiality, this is easier said than done in our polarized times.

A final worry that must be noted in this discussion of revenue regulation is that certain segments of the general population will likely be unable to participate in social media discourse if a transition away from the advertising revenue model is required by government regulation. It is an unfortunate reality that some people cannot afford various forms of media because they need to direct their limited funds towards their basic material needs. While it is possible to provide exceptions and carveouts for such individuals in order to avoid excluding them from social media platforms, it is nonetheless the case that making significant changes to the revenue model of social media companies will likely make participation more difficult or impossible for members of demographic groups who are already marginalized in a variety of ways. If we are interested in protecting the social good of equity in accountability, as well as other forms of equity, then this is a concern that must be confronted when contemplating revenue regulation and its ability to usher in new dynamics in the realm of social media discourse.

While I have been critical of amplification regulation and revenue regulation here, it does not follow from this that these techniques for addressing intimidation culture ought to be discarded. To the contrary, they ought to be analyzed and discussed further. If these regulatory approaches are formulated carefully, with adequate attention being paid to potential unintended consequences, then they may prove valuable assets in the fight to rein in intimidation culture and cultivate productive discourse across society. Accordingly, the view that state intervention into the social

---

into communities of like-minded types, their own freedom is at risk. They are living in a prison of their own design." (2018, 11-12)

media marketplace is inherently injurious to free expression ought to be rejected. Some well-crafted regulation in the social media sector may actually help to bolster free expression and the social goods that it helps to secure. However, at present, there is little reason to think that these regulatory techniques or others that have been proposed amount to comprehensive remedies to the pernicious phenomenon that motivates this discussion. They may indeed amount to steps in the right direction, but they fall short of realizing the ultimate objective of overcoming intimidation culture and supplanting it with a healthier culture of intellectual openness and diversity.

iv.iv: Voluntary Exodus from Social Media

After having considered government bans and government regulation of social media and offering an assessment of these strategies for addressing the phenomenon of online intimidation culture, it is appropriate to examine an alternative approach that does not invoke the power of the state. It might be argued that this strategy is the most straightforward of all the options that are explored in this chapter, making it especially appealing in terms of parsimony. This strategy simply involves consumers in the marketplace voluntarily engaging in an exodus from social media platforms, thereby rendering them irrelevant. Social media companies wield power and influence precisely because of their large userships, and if users abandon these platforms in droves, then it will surely be the case that the power and influence of social media over public discourse will be eroded. If people choose to escape from the dysfunctional realm of social media discourse in high enough numbers, and turn to alternative venues in order to express themselves and engage with others, then the problem of online intimidation culture will effectively evaporate.

Some may dismiss the notion that large swaths of the public will voluntarily part with social media. Indeed, an observation that animates this entire discussion in the first place is that social media is now intertwined with countless human affairs. These include interpersonal relationships, professional networking, journalism, the democratic process, commerce, and many others. Social media is not something that can be abandoned without serious consequences in a variety of areas, and it would be wrong to simply view this technology as a distracting toy that can be given up without any significant costs. With these caveats in place, it may be worthwhile to look to relatively recent history for an example of an industry that has experienced voluntary abandonment on behalf of consumers. Perhaps this can help us arrive at a sound assessment of the idea that online intimidation culture can be overcome via a voluntary exodus from social media platforms.

Cigarette smoking was once a ubiquitous practice, but it has seen an enormous drop in popularity as consumers have become more aware of its harmful effects.[133] While regulation has certainly played a role in this area, it is nonetheless true that the transition away from cigarette smoking has ultimately been the result of consumers becoming more informed and making increasingly conscientious decisions in the marketplace, rather than a result of strict rules being put in place by elites: while cigarette smoking has become more expensive and inconvenient as regulations have been implemented in various jurisdictions, it remains a legal activity that adults may partake in as much as they see fit. More recently, alcohol sales have experienced a decline, indicating that this market is vulnerable to changes in attitudes among consumers as well.[134] This phenomenon appears to be more grassroots in nature than the decline of cigarette smoking, as it

[133] See Christensen 2023.
[134] See Chong 2023.

has taken place in the absence of an aggressive effort on behalf of governments to discourage alcohol consumption.

When we take these trends into consideration, then the idea of a voluntary exodus from social media does appear more plausible. It is possible for behaviours that are extremely widespread to be phased out, even when sizeable industries stand to benefit financially from these behaviours. If consumers become increasingly convinced that social media use is damaging to their own wellbeing as well as the wellbeing of society, then perhaps we can expect them to increasingly reject social media services, just as they increasingly reject physical substances that they perceive as being harmful. The upshot is that the notion that online intimidation culture can be addressed through consumers voluntarily phasing out social media through their marketplace behaviour merits serious consideration, as consumers are entirely responsible for the relevance of these platforms in the first place and can put an end to this relevance via their consumption choices.

While it is true that the case for a voluntary exodus from social media should not be dismissed in light of trends that can be observed in other areas of the economy, it is nonetheless true that this strategy carries with it real disadvantages. The foremost concern with this strategy for mitigating social media's damaging impact on public discourse is that it does not account for the complex combination of benefits and drawbacks that are involved in social media use. Although we ought not shy away from criticizing social media platforms for the perverse incentives that they introduce into public discourse, it cannot be denied that people do benefit from social media use on a routine basis. In addition to being cacophonous and dysfunctional in many ways, social media platforms function as powerful tools for worthwhile projects such as career

advancement, forging and rekindling relationships, and bonding over shared interests. Even professional academics frequently look to social media in order to interact with a community of scholars, promote their own work, and obtain recommendations about worthwhile literature by others. While a voluntary exodus from social media could take place if consumers become fed up with these platforms, and it would indeed help to rein in the toxicity that pervades our discourse, abandonment of social media would also mean abandonment of the positive aspects of this technology. This is where an analogy between social media and other harmful commercial products, such as cigarettes, becomes problematic. It is very plausible to argue that cigarettes provide no meaningful benefits to their users, but it would be erroneous to say the same of social media platforms.

There is another reason to be concerned about calling for a voluntary exodus from social media. As stated in this chapter and the preceding chapter, one of the key social goods identified in this work is equity in accountability. I have argued that one of the aspects of online intimidation culture that makes it pernicious is that it is most stifling to segments of society that have the least in terms of economic privilege. An issue with calling for a voluntary exodus from social media is that this proposed exodus may exacerbate inequity in its own way. While abandoning social media may be relatively painless for people who already have an excellent career, an excellent social life, and excellent access to a broad array of media, abandonment will prove to be more painful for people who rely on social media platforms for the sake of generating income, keeping their social relationships alive, and accessing information about niche interests. When we seriously scrutinize the project of abandoning social media, we may find that this strategy is most congenial to those who are exceptionally well off, and most disadvantageous to those who are less fortunate. For

those who are less privileged, quitting social media may involve real material losses without any clear counterbalancing gains in the short term. We must take these types of concerns seriously before jumping to the conclusion that abandonment of social media is the appropriate path forward with respect to the alleviation of online toxicity and its damaging impact on public discourse.

Another noteworthy reason why a voluntary exodus from social media is problematic for our purposes concerns the dynamics of intimidation culture. It was argued in Chapter 1 that the personal attacks that are facilitated and encouraged by social media platforms are responsible for generating enormous chilling effects. People do not want to be on the wrong side of a social media firestorm, which causes them to tread very carefully in online and offline settings, and even to falsify their own views. A problem with the strategy of addressing intimidation culture by voluntarily exiting social media is that it removes a line of defence that is available to people when online attacks do occur. In certain cases, participation on social media enables people to correct the record, so to speak. For example, if a person is accused of being a brutal warmonger, and yet their social media profiles clearly contain numerous posts advocating for peace, then this provides a certain amount of inoculation from the damaging accusation being made. This inoculation may be insufficient, but it is still potentially meaningful. Meanwhile, if a person is entirely absent from social media discourse, they may find that they have no effective means of protecting themselves when personal attacks are launched in online settings.

The upshot is that while a voluntary exodus from social media would obviously succeed in addressing intimation culture if and when large portions of the population act in concert, it may be counterproductive for specific individuals and groups who are attempting to spearhead this exodus.

While it is commonsensical to think that exiting social media effectively shields an individual from online attacks, this is not true in all cases, and we ought to remain cognizant of this. Online intimidation culture may even prove more pernicious in cases wherein its target is not a participant in social media discourse, and accordingly has no direct means of deescalating a social media controversy. It would be an unfortunate irony if individuals started to abandon social media in hopes of remedying online intimidation and steering culture in a better direction, only to find that they have rendered themselves less equipped to handle online attacks if and when they do take place. While it is true that some people choose to exit social media and are happy with this decision, it is an error to think that exiting social media in any way guarantees freedom from the hostility and vitriol that dominate online discourse on a routine basis.

It might even be argued that a social media exodus effectively amounts to the ultimate triumph of online intimidation culture, and a defeat for those who wish to cultivate an atmosphere of free expression. While this discussion has repeatedly criticized social media companies and the dysfunction that they encourage, it has also been noted that there are many significant benefits afforded by this technology. There is a lot of good that can come from using social media and from engineering platforms that contain healthy incentives, and it would be a significant loss for society if social media were completely discarded. Abandoning social media for the sake of reining in intimidation culture would signal that people are not competent enough to address this issue without throwing away all of the positive features and experiences that are associated with this technology. In some sense, a widespread social media exodus amounts to simply giving up in the face of a serious challenge. This is obviously suboptimal given the fact that there is no logical reason why toxicity and intimidation must be permanent features of social media.

Before this discussion of a voluntary exodus from social media is concluded, a point of clarification is needed. There is a middle ground between the notion that consumers ought to continue using social media as they already do, and the notion that social media ought to be discarded. One may argue that social media platforms ought to be abandoned not permanently, but until the technology industry figures out how to deliver social media services to consumers that are more conducive to rational and constructive dialogue.[135] It has been established that social media platforms inject incentives into public discourse that are profoundly damaging, but it does not follow from this that such bad incentives are unalterable features of social media. It is possible for the incentives that permeate social media to be realigned so that users are no longer rewarded for participating in displays of outrage, hostility, and self-righteousness, creating space for more productive forms of discourse to flourish. Accordingly, it is logical for consumers to engage in a temporary exodus from social media that is intended to pressure companies to clean up their platforms, rather than to quit the platforms permanently on the grounds that they are damaging to individuals and society.

I, like many others, am fond of the idea that consumers can pressure companies into improving their conduct by abstaining from using products and services that carry significant harms. This is a much more optimistic outlook than the notion that social media ought to be abandoned permanently, and it is undeniable that social media companies are willing to revise

[135] Lanier argues that a promising strategy for introducing healthier incentives into the technology industry "is to directly monetize services such as search and social media. You'd pay a low monthly fee to use them, but if you contributed a lot—if your posts, videos, or whatever are popular—you could also earn some money. A large number of people, instead of the tiny number of token stars in the present system, would earn money. (I acknowledge, of course, that there would have to be a way of making services available to those who couldn't afford to pay even a small fee.)" (2018, 104)

their services in order to attract and retain users, as the user experiences of these platforms have evolved dramatically since they first came to prominence in the 2000s. However, it must be acknowledged that if we examine the history of social media, it is not clear that departure from a platform is a reliable mechanism for generating improvements on said platform. Indeed, the precise opposite can be the case. When a social media platform begins to experience an exodus of users, this entails losses of revenue and talent that can effectively steer the platform into a spiral of increasing dysfunction. While there is room for some debate about this matter, the disorder that took place at Myspace and Twitter when these companies entered a period of decline suggests that wounding a social media company via consumer exodus can bring about a user experience that is unambiguously worse than what preceded it as companies desperately try to keep their business afloat in the face of shrinking resources.[136] While it is possible for different companies to respond to fleeing users in various ways, this is a topic that should not be overlooked if our goal is to create a media ecosystem that is less prone to dysfunction than the one that has flourished in recent history.

In the preceding section, I argued that while some strategies for regulating social media via government intervention have merit, they also carry unintended consequences that are worthy of concern. A similar assessment seems appropriate here. There is no doubt that social media users have power in their relationships with social media companies, despite the fact that the latter possess far more resources and have much greater reach across the globe. Indeed, the ability of ordinary people to help foster a better information environment will be discussed in greater depth

[136] Jason Hannan offers a glimpse into the turmoil that unfolded at Twitter after its acquisition by Elon Musk: "In between his endless joking, jesting, teasing, mocking, prodding, and exuberant guffawing, Musk had to warn Twitter's surviving staff members that without sufficient revenue to stay afloat, bankruptcy was not out of the question." (2023, 125)

in the chapters that follow. However, the possibility of a voluntary exodus from social media raises significant concerns even if we accept the premise that the exodus should only be temporary. While we should always remain open to the idea that some products and services are so toxic that people are simply better off without them, it would be wrong to assume that abandonment of social media will bring about the desired results in a manner that is timely, efficient, or predictable. Abandonment of social media can involve significant costs for those who choose to engage in this exodus, and it can also backfire by introducing even greater instability and dysfunction into the media ecosystem. Accordingly, if we are to identify a sound and efficacious means of overcoming online intimidation culture, it is appropriate to embrace a broader view of users' relationship with social media and how it can be improved.

iv.v: The Absence of a Panacea

This chapter has offered an overview of strategies for addressing the phenomenon of online intimidation culture. I have argued that while a government ban of social media would obviously realize the objective of eliminating this technology's ability to shape expression in pernicious ways, it would raise enormous concerns about free expression and the stifling of dissent. While certain influential individuals are in favour of banning specific social media platforms, the notion that social media itself ought to be banned in a wholesale manner remains extreme and impractical, and is unlikely to gain a foothold in mainstream culture anytime soon. A more measured approach that is far more likely to achieve widespread support involves regulating social media companies via the state, and seeking to realign the incentives that are present in online discourse. While I have noted that the strategies of amplification regulation and revenue regulation show some significant promise and are worthy of serious discussion, they do not provide a comprehensive solution to the

problem that motivates this discussion. If amplification regulation and revenue regulation are going to be pursued in the policy arena, they ought to be conceptualized as components of a broader plan to grapple with the problems generated by social media. Neither strategy has a realistic prospect of solving the problems that motivate this discussion. It has also been argued that while a voluntary exodus from social media could significantly pressure social media companies into revising their conduct, this strategy too raises a host of concerns, including ones that are directly connected to the dynamics of online intimidation. Abandonment of social media remains a risky strategy for those who wish to combat its pervasive chilling effects and play a role in fostering a healthier information environment.

As we can see, the strategies that have been outlined in this chapter vary considerably in terms of their overall plausibility; some are far more practical than others. The conclusion that is appropriate in light of the arguments presented above is that while there are some promising strategies available with respect to the issue of online intimidation culture, at present there is no straightforward panacea that we can readily access. Online intimidation culture will not be eliminated swiftly: while it can be tempting to think that our societies are one clever policy fix away from remedying the pernicious impact of social media on public discourse, the complexity of this issue ought to temper our expectations in this area. Intimidation culture is an impressively vast phenomenon that implicates many facets of human life. Accordingly, the chapter that follows will make the case that the process of building resilience to online attacks at the institutional level can help rein in the influence of online toxicity in the interest of protecting free expression and the social goods that it produces. Since social media has played a pivotal role in derailing public

discourse and the norms that surround it, we will now explore the topic of how our institutions might be amended in order to steer public discourse towards a healthier trajectory.

Chapter v: Cultivating Resilient Institutions

v.i: Toxic Media and Cultural Antibodies

The preceding chapter provided an overview of some strategies for mitigating the pernicious effects of social media platforms. It has been argued that none of these strategies amount to a panacea. An outright ban of social media would be overwhelmingly illiberal, and would raise concerns about free expression that far outweigh the reasonable concerns that surround the issue of online intimidation culture. While regulating social media companies and voluntarily exiting social media can have some significant and beneficial outcomes, there is little reason to think that such strategies are equipped to eliminate the pernicious dynamics that pervade our contemporary media ecosystem. We will now turn our attention to institutions outside of the social media sector in hopes of identifying a strategy for addressing the phenomenon of online intimidation culture that is both highly efficacious and aligned with Millian, liberal objectives. It might turn out to be the case that by reforming a variety of institutions that are not directly connected to the social media industry, we can limit the ability of social media companies to derail public discourse and undermine freedom of expression.

A key premise that animates this chapter is that it is possible, and desirable, for societies to build resilience with respect to forms of communication and behaviour that are decidedly toxic. It is obviously the case that whenever a form of communication or behaviour is legitimately deemed to be destructive, a normal reaction is to want to purge it from society. This applies in the domain of social media just as it applies in countless other domains. It is natural for those who analyze social media and the overwhelming hostility and outrage that it facilitates to want to put a stop to these pernicious patterns of communication through various means, including government regulation, boycotts, and design changes to social media platforms. This is a noble goal. However,

in the absence of a straightforward mechanism for achieving it, we will need to broaden our perspective in order to address the problem of online intimidation culture.

An analogy with the domain of physical health can be helpful for understanding how we can grapple with intimidation culture without (directly) eliminating toxic discourse from social media. We generally have a good understanding of the idea that while it would be ideal to live in an environment wherein antigens that threaten our wellbeing are absent, this is not realistic most of the time. We live in a world that is teeming with substances that can harm us, and while we may successfully limit or control these substances part of the time, we do not have the power to eliminate them completely. Accordingly, if we are interested in securing our wellbeing over the long term, we must invest in the project of developing antibodies that can protect us from the many antigens that have the power to cause us harm. While antibodies cannot shield us from every illness or injury, they are an enormously valuable resource, and we would be in deep trouble without them. Antibodies do not remove damaging substances from our environment, but they do assist us in coping with these substances and ensuring that our lives are not derailed when we come into contact with them.

My contention is that the reasoning that we deploy regarding the importance of immunity in the context of physical health ought to be extended into the realm of media. It is clear enough that social media can be toxic in a variety of ways, and I obviously share this concern with many other scholars and commentators. We can all agree that society would benefit from a media ecosystem that is less toxic, and there is no reason to give up on the goal of making this a reality. However, a question that has received insufficient attention is how individuals and institutions can

respond to this toxicity given that it is not feasible to cut it off at its source. If we are serious about addressing the issue of online intimidation culture then it is appropriate to be realistic and accept that the vitriolic character of social media discourse is not going to disappear immediately. It is an unfortunate reality that online discourse continues to be intemperate and divisive, and that social media use continues to be a staple of everyday life for vast portions of our planet's population. With this being the case, we can put ourselves in a position to remedy the pernicious influence of toxic online communication by directing our attention not merely towards social media discourse itself, but also the array of institutions that are negatively impacted by social media toxicity, as well as the individuals that enable these institutions to operate.

This idea can be phrased another way. Instead of asking how the toxicity of social media discourse can be curtailed, we can instead ask how to make our society more resilient in the face of this toxicity. If the ability of online toxicity to inflict damage on society is progressively diminished, we can refer to this process as one wherein "cultural antibodies" are produced. Just as literal antibodies can protect our health in a world that is pervaded by various antigens, cultural antibodies can protect the health of our public discourse in a world that is pervaded by toxic communication on social media. While we should not lose sight of the project of reforming existing social media platforms and constructing new platforms that are more conducive to good-faith dialogue, there is no reason why our ambitions should stop there. We can also work to address the phenomenon of online intimidation culture by striving to cultivate a society that is more resilient in the face of rampant online attacks. Moreover, if we succeed in cultivating societies that are more resilient to social media attacks than they are at present, this may have downstream benefits that are not immediately apparent. This chapter posits that the development of resilience in institutions

and individuals amounts to a potent strategy for reining in the problem of online intimidation culture and reducing its corrosive impact on free expression and the social goods associated with free expression. Importantly, this is a strategy that is highly congruent with the Millian aspirations that were outlined in Chapter 2, and entails far fewer social costs than the strategies that were explored in Chapter 4.

v.ii: The Project of Hardening Institutions

When discussions about social media make references to institutions, the institutions that are usually invoked are governments as well as social media companies themselves. It is not difficult to see why this is the case. No institutions are more intimately connected to the issue of social media discourse than the companies that create and maintain social media platforms, and these companies exert enormous influence over the tone and character of the conversations that unfold in online settings. Meanwhile, governments are expected to carry out the will of their citizens - at least to a certain extent - so it is unsurprising that critics of social media would seek to deploy government power when it becomes evident that the dynamics of social media discourse are decidedly at odds with democratic values and the public interest. Previous chapters have supported the notion that social media companies and governments ought to be thoroughly scrutinized for their involvement in the perpetuation of online intimidation culture, and nothing in this chapter challenges this notion.

However, it is now appropriate to take a broader view of how institutions can respond to this worrying phenomenon. Rather than focusing exclusively on governments and social media companies, we ought to examine how other important cultural institutions can go about reining in

online intimidation even when they are not directly implicated in social media discourse. The overarching point is that these institutions can help society develop resilience in the face of rampant online hostility by embracing policies and protocols that acknowledge the problem of online intimidation culture and seek to address intimidation campaigns before they take place. This is a strategy for coping with online toxicity that clearly transcends the purview of governments and social media companies, and accordingly can enable societal actors who are concerned about the pernicious impact of social media on public discourse to bypass the cumbersome process of wrangling with these institutions, and to effect positive change with respect to discourse in other ways.[137] This approach may be more streamlined than other approaches that focus specifically on governments and social media companies, meaning that it can operate and produce results in a more swift and efficient fashion.

Perhaps the clearest and most compelling case for an organized, institutional response to the pernicious phenomenon of online intimidation culture comes from Brookings Institution Senior Fellow Jonathan Rauch. He states the following about the subject:

> Most important of all is for employers and companies to internalize resistance to cancelations, especially by preparing for attacks. In order to defend their values when a crisis hits, organizations need to identify and declare their values ahead of time; otherwise, they panic and cave in. They can prepare by (for example) setting up internal procedures preventing a rush to judgment against targeted employees; by pre-committing to evaluate the totality of an employee's work history and character rather than acting on the basis of a single controversial action or allegation; by offering recourse and support to employees who are targeted on social media (or by bullies inside the company); by promulgating

[137] The following was circulated in 2023 by *The New York Times* after some of the newspaper's own personnel engaged in online castigation in order to shape the work of other NYT journalists: "It is not unusual for outside groups to critique our coverage or to rally supporters to influence our journalism. In this case, however, members of staff and contributors to The Times joined the effort. Their protest letter included direct attacks on several colleagues, singling them out by name. Participation in such a campaign is against the letter and spirit of our ethics policy … We do not welcome, and will not tolerate, participation by Times journalists in protests organized by advocacy groups or attacks on colleagues on social media and other public forums." This statement suggests that institutions are increasingly aware of the ability of social media to chill discourse and drive conformity, and are taking firmer stances in order to rein in peer pressure and intimidation.

> guidelines for human resources and communications executives to follow when cancel campaigns boil up; and, above all, by expressing a commitment to their employees' off-workplace speech rights. By hardening their defenses, organizations make themselves more resilient if hit by cancelers—and therefore less tempting as targets. (2021, 239)

Rauch's reference to institutions "hardening their defenses" is helpful because it clearly conveys the idea that in lieu of doing away with the rampant hostility that pervades social media discourse, various facets of society can become stronger in the face of it. I strongly agree with the notion that by shielding their personnel from the most severe consequences of targeted online attacks, institutions can help alleviate chilling effects and generate an intellectual climate that is more tolerant and conducive to heterodoxy. While I have accepted the prominent view that the toxicity of social media is producing social censorship and self-censorship across society, there is little reason to think this toxicity must invariably carry so much heft and influence. If an array of cultural institutions make a concerted effort to strengthen their defenses against the belligerent actors that take up so much space in the realm of social media, this will undermine the power of these belligerent actors to shape society in pernicious ways. Instead of bringing about the end of online vitriol, institutions can adapt to the new reality of omnipresent social media toxicity and take steps to reduce its impact to that of a background noise that has little, if any, ability to steer the direction of society. If we cannot get rid of the cacophony of social media, then we ought to take steps to contain it and mitigate its societal effects. This project would remove power from actors that wish to use the Internet in order to frighten dissenters and pressure them into conformity, and place this power with more responsible actors who wish to present ideas in an open and tolerant manner so that these ideas can be rigorously analyzed and assessed.[138]

[138] Mounk offers a similar argument: "Anybody who cares about upholding a genuine culture of free speech must … care about reining in the ability of private actors to punish people for expressing unpopular views or to police the boundaries of legitimate debate. Thankfully, governments can help to constrain private power without overstepping the strict limits on what they themselves can legitimately do in this realm. The first step should be to ban companies

While there are many facets of society that are shaped by online intimidation culture to some degree, it is reasonable to suggest that those that are directly implicated in the generation and dissemination of knowledge deserve special attention when it comes to the project of hardening institutions. Academia, journalism, and publishing stand out as areas wherein an organized response to intimidation culture is appropriate and likely to produce desirable effects with respect to the health of public discourse.[139] Since people turn to these institutions specifically for their ability to provide valuable information rather than their ability to produce more mundane goods, it is logical for them to take firm positions with respect to intellectual diversity and the corrosive impact that social media can have on it. While public discourse certainly transcends institutions in the areas of academia, journalism, and publishing, it is nonetheless the case that these institutions enjoy a level of esteem that many others do not, and can accordingly function as leaders with respect to modeling productive discourse and the tolerance for dissenting views that it entails. We may find that making institutions in these areas more resilient with respect to social media toxicity will involve spillover effects that inspire other institutions to follow suit. These institutions can help spearhead a cultural shift wherein online intimidation culture is increasingly understood as a force that can and ought to be marginalized through organized institutional responses.

from firing their employees for saying unpopular things. Governments could accomplish this by including the political views of employees in the list of protected characteristics, as some jurisdictions including Seattle and Washington, D.C., have already done." (2023, 177)

[139] Redstone and Villasenor provide discussions of social media's influence in all three of these domains. They argue that social media has stifled discourse in the realm of academia: "… social media have changed how we communicate and have emerged as a powerful tool both for direct censorship and for strengthening the incentives for self-censorship … Both on campus and off, this is most visible through social media–driven public shaming campaigns launched in response to perceived transgressions …." (2020, 2) They explain how social media backlash can influence mainstream news media in pernicious ways: " … call-out campaigns that start on social media can quickly cross over to mainstream news media, and in both settings the story is largely defined by those who raise the loudest criticism—even if they don't necessarily represent a majority view." (2020, 36-37) The authors state the following about the case involving young adult fiction author Kosoko Jackson: "... it illustrates the power of a social media mob to exercise veto power over publication decisions that should more properly be made by publishers, editors, and authors." (2020, 140)

It is worth noting that the kind of policies for which I am advocating, alongside Rauch, are not without precedent. The Kalven Report was written in 1967, and encourages universities to embrace institutional neutrality with respect to hot-button social and political topics in order to ensure that these places of higher learning maintain an atmosphere of intellectual freedom and tolerance of dissent. The report offers the following declaration:

> The neutrality of the university as an institution arises … not from a lack of courage nor out of indifference and insensitivity. It arises out of respect for free inquiry and the obligation to cherish a diversity of viewpoints. And this neutrality as an institution has its complement in the fullest freedom for its faculty and students as individuals to participate in political action and social protest. It finds its complement, too, in the obligation of the university to provide a forum for the most searching and candid discussion of public issues.[140]

Redstone and Villasenor invoke the Kalven Report in order to underscore the importance of academic departments and organizations remaining hospitable to a variety of perspectives:

> Academic freedom can … be undermined through the implicit silencing of voices that express views that are out of step with the views of a majority of people in an academic department or an academic organization such as a scholarly society. This is apparent through the numerous examples where faculty have used departments and academic organizations as platforms to make statements on political issues. It also runs counter to recommendations made by the Kalven Committee … The *Kalven Committee Report* (also known by the more formal title *Report on the University's Role in Political and Social Action*) was issued by the University of Chicago after the committee was convened by the university president to prepare 'a statement on the University's role in political and social action.' (2020, 55-56)

As we can see, the Kalven Report does not say anything to discourage individual faculty members, students, or administrators from voicing their own views about controversial issues as forcefully as they see fit.[141] It simply recommends that universities avoid taking official

[140] See Banout and Ginsburg 2024, 165.

[141] Keith E. Whittington argues that intellectual diversity is key to the "institutional health" of universities: "Professors are hardly unique in being exposed to the ill humors of Internet mobs or discovering that an incautious post on social media has angered their employers … It should be understood that individual faculty members do not speak for or represent the institution. Rather, the institution houses dozens or hundreds of diverse and conflicting faculty members. The 'brand' to be protected in the case of the university should be the one reflected in its institutional mission of facilitating the pursuit of knowledge through vigorous debate and open inquiry. The presence of unorthodox, controversial, and even wild-eyed professors on the faculty should be regarded as a sign of institutional health. The

institutional positions on such matters, thereby signalling that members of the university community who are not aligned with the university's institutional position are unwelcome or illegitimate. The Kalven Report is prophylactic in the sense that it strives to prevent institutions from degenerating into intellectual monocultures that lack dynamism and openness, and its prescriptions seem entirely appropriate with respect to achieving this objective. Organized institutional responses to online intimidation culture can be understood as similar mechanisms that seek to prevent institutions from being corrupted by the enormous social pressure that social media platforms facilitate. These policies are Millian in nature, and can be beneficial with respect to the internal operations of institutions, helping to cultivate meaningful and productive discourse, while also being beneficial with respect to the manner in which these institutions are viewed and perceived by individuals and groups that are external to them.

Having noted that academia, journalism, and publishing stand out as key areas wherein institutions can play meaningful roles in combating the influence of online intimidation culture, it is appropriate to note that journalism deserves particular emphasis and scrutiny because of its role in incentivizing and fuelling this phenomenon. Over the course of the 2010s, it became common for journalistic outlets to report on online controversies and disputes, thereby increasing their reach and encouraging greater participation in them. While it would be foolish to suggest that journalists ought to refrain entirely from covering social media spats, it is perfectly reasonable to posit that our media ecosystem will be better off if journalistic outlets exercise greater care and conscientiousness when deciding whether to amplify online skirmishes. Public policy researcher Renee DiResta has the following to say about this matter:

---

far larger threat to the reputation of a university should be the stifling docility of 'cautious mediocrity' or the unimaginative regimentation of ideological conformity." (2019, 153-154)

> 'Think before you share' is a simple rule that could dramatically transform the information environment today; however, it's not only individual consumers who need to learn that lesson. Media, too, amplify the most sensational trends on social media, often covering absolute nonsense pushed by a relatively small handful of people: 'Some people on the internet are saying . . .' Media coverage of a small, sensational controversy can amplify its reach significantly. Researchers like FirstDraft work on educating reporters to be aware of manipulation tactics; those at Data & Society have urged reporters to practice 'strategic silence'—that is, choosing not to cover (and thus amplify) speech known to be false … perhaps it's time to update that playbook to include other forms of speech that … are simply designed to foment outrage and generate clicks. (2022, 136, in Bollinger and Stone, eds.)

It has been noted that these battles often accomplish very little besides increasing engagement and perpetuating hostility, and so in many cases it is inappropriate and unhelpful for journalistic outlets to fuel them via coverage. In the case of journalism, hardening institutions will involve not only embracing policies that shield personnel from online attacks, but also discouraging personnel from spotlighting online attacks in a manner that is pernicious with respect to the health of discourse.[142]

The project of hardening institutions is connected to the project of promoting social media literacy. If journalistic institutions dedicate time and effort towards training their personnel on how to use social media in a responsible manner and avoid fuelling unproductive online quarrelling, this will generate greater social media literacy among journalists. If institutions codify these policies and publicize them this, in turn, will generate greater social media literacy among the public as they come to appreciate how journalistic media and social media can influence one another in healthy or unhealthy ways. While it is obviously the case that mega institutions such as federal governments and social media companies can play a role in cultivating social media

[142] Stone notes: "When the media presents exaggerated evidence of polarization and excessively negative representations of out-partisans, this leads to excessive perceived polarization (false polarization) and affective polarization bias because we fail to account for the way the 'sample' is selected. It's indeed hard to understand this selection process … Even journalists appear subject to selection neglect; for example, they treat Twitter users as a more representative sample of public opinion than they really are …" (2023, 141)

literacy, there is no reason why we must wait for these behemoths to act. Smaller sites of power such as journalistic institutions can play a significant role in driving productive cultural evolution via social media literacy, and they can do this in a relatively quick and efficient manner.

Educational institutions of course stand out as organizations that can play a role in advancing social media literacy. While it may be ideal for social media literacy to be incorporated into educational curricula in an official manner, educators can take action in this area in the absence of such formal reforms. Classes about subjects such as business, computer technology, communications technology, and psychology can function as gateways towards productive discussions about social media's societal impacts and how students can resist participation in toxic online dynamics. Moreover, since it is increasingly common for schools to establish rules regarding social media use in their official codes of conduct, these rules can be supplemented with materials that clearly convey how social media incentivizes bad behaviour. Once students are made aware of these incentives and the damaging behavioural patterns that they encourage, they will be better equipped to develop healthy relationships with social media. One might even argue that in order for rules about social media use to be legitimate, educational institutions must establish an adequate baseline of understanding among students regarding how this technology works and how their online conduct can impact others.

The introductory chapter of this work made the case that the dynamics of social media discourse are generating immense chilling effects. This remains the primary concern that animates this discussion. The project of hardening institutions is promising and worthwhile because it has the potential to create "warming effects" that can counteract the chilling effects that have been

documented over the course of the 2010s and 2020s. If we cannot easily or straightforwardly change the tone and character of social media discourse, it is incumbent upon us to explore other strategies for addressing the problem of online intimidation culture. Hardening institutions with respect to the issue of online intimidation is desirable because it has a strong chance of producing positive results, and also because it does not prohibit us from pursuing progress in other areas. Engaging in the project of hardening institutions does not exclude one from pressuring social media companies into changing their content moderation policies or from encouraging governments to embrace sensible regulations that can bring about change with respect to the issue of online intimidation culture. Accordingly, it is a project that ought to be embraced and championed by those who are concerned about social media and its corrosive impact on free expression and the social goods that free expression helps to secure.

v.iii: Institutions and the Realignment of Incentives

The project of hardening institutions is worthwhile from a Millian perspective because it can help realign incentives that are at play in public discourse in helpful ways. For example, if an employer adopts a policy that states that in cases of online mobbing, no targeted employee will be reprimanded or fired before a substantial period of time has elapsed, this will send the message to employees that they have strong institutional support on their side.[143] They will accordingly be less fearful of online attacks being ruinous to their career and livelihood, and be incentivized to express themselves in an authentic manner rather than to tailor their speech so as to avoid upsetting

[143] Mounk writes that universities should make efforts to reduce chilling effects: "The core purpose of universities … is to produce knowledge. Given how easily that purpose is subverted by social pressure or the fear of being fired, they should voluntarily adopt strong protections for 'academic freedom' (as many of them have, at least on paper)." (2023, 176)

the partisans and ideologues that dominate social media discourse.[144] This would be beneficial with respect to the health of discourse, especially if we accept the idea that social media platforms have been a boon to the more extreme individuals and groups in society, enabling them to exert outsized influence over the rest of the population.

With that being said, it is also important to note that hardening institutions can have desirable effects with respect to the incentives that are at play on the other side of social media controversies. This is a point that was alluded to in the quotation from Rauch that was included above, and I wish to say more about it here. Hardening institutions is valuable because in addition to sending a signal to individuals inside of institutions, it also sends a signal to external agitators who would like to inflict reputational and professional damage upon an institution's personnel. If institutions begin confronting the problem of online intimidation culture, this will convey to social media users that launching online attacks on individuals and groups that they dislike is unlikely to have its intended effect. This disincentivizes social media users from launching these attacks in the first place, which is clearly valuable if we are interested in steering social media discourse in a direction that is less accusatory and combative, and more understanding and constructive.

Some may question the value of realigning incentives in this way on the grounds that what social media combatants are truly after is not to inflict reputational and professional damage on

[144] Tosi and Warmke argue that social media can be used to exert dominance over others: "Dominance … refers to the status you get by instilling fear in others through intimidation, coercion, or even displays of brute force. The dominated treat you with deference because they fear being treated harshly … In modern times, people still use physical violence, but we can also gain dominance by embarrassing others on social media, or lashing out at a colleague in a meeting." (2020, 16-17)

their adversaries, but rather to garner approval and social status among members of their own ideological camp.[145] Technologist Tobias Rose-Stockwell details this dynamic:

> Social media is built on the back of many innate human impulses. The fundamental desire for prestige and social status underpins our desire for followers and likes, for example. Cancellations and callouts are a function of the human impulse to shame others when they transgress the moral norms of society. It's an amplification of a core human social behavior: gossip. .... Indulging in collective schadenfreude is an oddly pleasant experience: a small retweet, a like, a share, a repost… it all seems innocuous, simple joy. A tiny guilty pleasure that brightens your day. But each of these is a tiny vote for condemnation of the recipient. And each of these votes can tally up to a life-changing stream of admonishment for the target. Having the passive ire of one million people directed at you is a debilitating sensation, one in which your life can grind to a halt. (2023, 162-163)

This commentary does highlight a real limitation that is involved in the project of hardening institutions. Even if institutions stand strong in the face of online attacks towards their personnel, the people responsible for launching these attacks can still enjoy other kinds of rewards such as increased online engagement and approval from their peers.[146] While derailing the life of a targeted individual might be an ideal outcome among those who are fond of participating in intimidation campaigns, it need not be achieved in order for social media users to experience group approbation and the emotional satisfaction that comes with it. As we know, many academics and technology experts who study online antagonism have noted that people who engage in aggressive online behaviour often appear to be largely motivated by the goal of cementing their own social status.

[145] Rauch articulates this point: "When I join others in a shaming campaign against you and bomb your Twitter account with imprecations, my tweets may take the form of communications to you, but in fact they are about you—and, especially, about me. What I am really doing is trying to impress my peer group with my virtue, cleverness, and loyalty. By joining the shaming campaign, and better yet by leading it, I can raise my status. You have the misfortune of being a useful object in my quest." (2021, 128)

[146] Nancy L. Rosenblum and Russell Muirhead highlight the role of "social validation" in perpetuating the conspiracism that pervades the digital age: "What validates the new conspiracism is not evidence but repetition … Forwarding, reposting, retweeting, and 'liking': these are how doubts are instilled and accusations are validated in the new media. The new conspiracism—all accusation, no evidence—substitutes social validation for scientific validation: if a lot of people are saying it, to use Trump's signature phrase, then it is true enough." (2020, 3)

This suggests that effecting tangible change in the real world is often a secondary objective for these individuals and groups, and that the development of social ties remains paramount.[147]

Hardening institutions and protecting individuals from personal and professional ruin in the wake of online attacks will not eliminate the social rewards that often accompany the circulation of such attacks. However, we do not need to eliminate these online dynamics in order to nurture an atmosphere of free expression and its attendant social goods. This is the overarching objective that animates this discussion, and it is where our focus ought to remain. Indeed, there are many segments of society wherein people enjoy social rewards for engaging in behaviour that can reasonably be viewed as pernicious. For example, advocates for outlandish conspiracy theories often join forces with other conspiracy theorists and bond over their shared interests and ideas, and encourage one another to delve more deeply into conspiratorial thinking. While this phenomenon may be concerning in its own right, it is reasonable to be truly alarmed only if and when these people begin to exert significant influence over the rest of society. If their ideas continue to be passed around relatively insular communities that are marginal and do not make other segments of society worse off, then this is a dynamic that we must be prepared to tolerate in a liberal society, at least to a certain extent. For the purposes of this discussion, it is appropriate to focus specifically on online social dynamics that threaten free expression, rather than to take on the broader and more difficult project of remedying unhealthy online social dynamics of all kinds. While the project of hardening institutions cannot eliminate all of the incentives that make social media discourse hostile and divisive, it can help confine this toxicity to relatively obscure corners of the Internet

[147] Vaidhyanathan emphasizes the importance of group affinity: "Social media make it easier than ever to identify the like-minded or potentially like-minded and hold them closely, giving them social rewards for adopting certain beliefs and fact claims. In this way Facebook simultaneously amplifies movements that use strong emotion to undermine trust in institutions and recruits and indoctrinates new believers in what used to be marginal beliefs." (2021, 16-17)

and prevent it from undermining the integrity of institutions as well as public discourse more broadly.

v.iv: The Dual Nature of Engagement: How Users Shape Social Media

A major theme that has informed this dissertation is the role of incentives in promoting and perpetuating bad behaviour in the realm of social media. We have seen that individuals and groups that are exceptionally self-righteous and divisive dominate social media discourse because this type of conduct is rewarded by the platforms themselves. Individuals and groups that strive to take a measured approach to complex issues, and engage fairly with interlocutors, struggle to gain a foothold in a social media landscape wherein emotionally-charged content is amplified via engagement and curatorial algorithms. This dynamic is what makes social media discourse far more toxic and less productive than other types of discourse, and it is also a feature of social media discourse that makes it a legitimate threat to free expression from a Millian perspective. Since social media discourse has the power to fuel social punishment and its associated chilling effects, it is important to think carefully about how this problem can effectively be remedied. I have accordingly argued that a process of hardening institutions can help rein in the pernicious impact of social media without doing away with this powerful and ever-changing form of technology.

At this juncture, it is appropriate to directly address a facet of social media that has only been briefly alluded to throughout the preceding discussion. While it is true that social media platforms have the ability to shape the behaviour of users through the incentives that they put in place, it is also possible for social media users to shape the behaviour of social media companies

through their online behaviour.[148] Just as social media companies can encourage or discourage the circulation of certain types of content by boosting or de-boosting it, social media users can do the same by choosing to engage with content or refraining from engagement. My contention is that since a good deal of attention and energy has been channeled by academics and other commentators towards the goal of holding companies accountable for their role in perpetuating online toxicity and corroding public discourse, it is reasonable to do the same with respect to users. This may prove valuable for the project of confronting and overcoming online intimidation culture and restoring the health of public discourse.

Institutions can go a long way in cultivating resilience to online intimidation efforts by educating individuals about the role that they play in sustaining and perpetuating online toxicity through their everyday actions. It was argued above that journalists deserve special attention when it comes to this issue, as they have the ability to significantly amplify content that is corrosive to the quality of dialogue. If journalists can become more conscientious with respect to the manner in which they approach incendiary and divisive content, so too can ordinary members of the public.[149] Since it is now almost universally the case for institutions to have a significant presence on social media, they ought to dedicate a portion of their resources towards educating their personnel about the economic and social incentives that are at play in these online platforms. Once people can clearly understand these incentives and their ability to warp discourse, they will likely

[148] While this discussion is focused on social media's role in facilitating intimidation, search engines also play a significant role. For a discussion of Google and its ability to shape people's reputations and life prospects, see Chapter 11 of Jon Ronson's book *So You've Been Publicly Shamed*. (2015)

[149] Mike Caulfield and Sam Wineburg note that bad actors benefit when Internet users rush to judgment about online content. The simple act of pausing and allowing for facts to be established can play a significant role in reducing the influence of bad actors: "Reliable reporters need time to work. Rage merchants do not. When it comes to breaking events, the greatest information literacy superpower is often just learning to wait before allowing yourself to form deep beliefs about the event. Remember that both the con artist and the propagandist feed on the impatient, because time to investigate and reflect does not favor liars." (2023, 164)

develop a more mature view of social media and be better equipped to engage in online communications without participating in patterns of behaviour that are damaging to individuals and society.

It is important to note that we are not talking about highly technical training here. We are talking about basic awareness about how the design of social media content can help or hinder its ability to ascend the social media timelines of users in order to broaden its reach and enrich the people responsible for creating it. For example, emotional language increases the reach of content, as does the inclusion of familiar human faces and provocative text in video thumbnails. While most institutions may not be in a position to generate major change at the level of government or social media company policy, they are certainly in a position to help people better understand how modern online platforms function so that they can conduct themselves in a manner that best reflects their own interests and the broader interests of society. [150] This decentralized process can be valuable for promoting social media literacy and disarming the noxious content that too often takes up the spotlight in the realm of social media, and crowds out other types of content that has a higher probability of generating productive dialogue. If undertaken, these efforts on behalf of institutions will likely function in support of their own interests to a significant extent by decreasing the likelihood that their personnel will become embroiled in online controversies.

[150] Caulfield and Wineburg argue that modest interventions can improve information literacy: "To date, thirteen separate studies involving nearly ten thousand participants have shown the effectiveness of our approach in helping people make better choices online. And in one of the most recent studies, students showed a sixfold increase in use of fact-checking techniques and a fivefold increase in citations of appropriate context after only seven hours of instruction." (2023, 5-6)

For better or worse, we live in an era wherein the online activity of individuals is frequently viewed as representative of institutions in which they are participants. Online attacks frequently call for people to be dismissed by their employers or educational institutions for this reason. These attacks tacitly advance the idea that if institutions fail to sever ties with individuals who are involved in online controversy, this amounts to an endorsement of their conduct. While fear of social punishment can and does shape the behaviour of individuals, it can also function similarly at the level of institutions. In this environment, institutions can benefit from explicitly acknowledging the fact that social media has the ability to inflict damage on them as well as the individuals that participate in their operations on a daily basis. Discussion of the incentives that are at play on social media can help institutions and their personnel avoid and de-escalate online battles before more dramatic measures become necessary. Alongside dedicating time and effort towards the advancement of social media literacy out of a general commitment to the public good, institutions can engage in such efforts for the sake of maintaining smooth operations and avoiding the turmoil that can take place when social media outrage is directed towards them and their personnel. Instead of addressing issues related to social media in a post hoc manner, institutions can reap gains by addressing them as part of their routine activities.

The positive implications of such institutional efforts are clear. If social media users become less vulnerable to the allure of extreme and incendiary online content, the ability of this content to dominate social media platforms and derail public discourse will be curbed. One does not need to be a professional economist to understand the basic dynamics of supply and demand in a marketplace. If demand for a specific type of content decreases over time, we can expect the supply to decrease as well, as actors in the media marketplace revise their conduct in response to

its signals and pressures. Since it is now commonplace for all kinds of individuals and groups to lament the bitterness that permeates social media platforms, it is reasonable to anticipate changes in the marketplace over the coming years and decades as consumers seek out different types of content that are less psychologically and emotionally draining than the content that has soared in visibility and popularity over the course of the 2010s and 2020s. Such changes are desirable, and they can be encouraged through organized institutional planning.

Ideally, greater awareness of the pitfalls associated with engaging with incendiary online content will help members of the public act more strategically so that they can help promote and generate content that is more prosocial rather than antisocial in character.[151] It was argued in Chapter 1 that engagement can plausibly be conceptualized as the currency of social media. Engagement, more than anything else, determines whether a piece of online content will reach a large audience or whether it will remain obscure. With this being the case, it is sound for people to work towards limiting their engagement to pieces of online content that they think have genuine merit and deserve to be promoted to a broader audience. While it may be tempting to share a piece of shocking or offensive content in order to express our disapproval of it, it is important to ask whether this is a wise use of the online tools at our disposal. If people become more savvy and conscientious with respect to their engagement patterns on social media, this will help reshape the user experience of social media platforms as companies learn that spotlighting extreme and emotionally charged content will fail to maximize attention and interest in the manner they seek. Engagement is dual in nature, enabling users to influence social media companies in addition to

[151] Aral explains that social media feedback can drive both detrimental and beneficial behaviour: "… the parts of our kids' brains that warn them that a behavior may be risky are turned off, or all least turned down, when photos depicting those behaviors receive more likes. … Positive social feedback encourages prosocial behaviors just as it does risky behaviors." (2021, 161-162)

being influenced by them. Since these companies want to be as relevant and profitable as possible, they must remain sensitive to the preferences of consumers in the marketplace and respond to the signals that they send out via their online behaviour. If companies fail to respond to the increasingly healthy preferences of consumers, they will be swept into irrelevance like many before them.

Much of the above discussion has emphasized the ability of institutions to highlight and remedy the problem of online intimidation culture. I have argued that hardening institutions is a promising strategy for increasing society's resilience to the toxicity of social media. While it is natural to want to eliminate such toxicity from the Internet in a sweeping manner, due to the fact that this is not going to be achieved overnight, it is incumbent upon us to participate in the development of cultural antibodies that can prevent toxic social media dynamics from wreaking havoc on society and the social goods that we wish to nurture and protect. At this juncture, it is appropriate to clarify that the project of cultivating resilience in institutions ought to be conceptualized as complementary to the project of cultivating resilience among individuals. The former is not a substitute for the latter. Rather, it is a gateway towards it. A society wherein institutions are committed to battling the corrosive effects of intimidation will be one wherein individuals are more likely to do the same.

If institutions remain steadfast in the face of online attacks and do not permit them to exert undue influence over their daily operations, individuals can be expected to behave in a similar manner. We might even say that institutional resilience can be conducive to the development of courage among a populace more broadly. Mill clearly endorses the view that nonconformity is

valuable in part because it signals to others that it is possible to resist social pressure and to develop a character that is not easily swayed by popular opinion. Since courage involves a willingness to take up unpopular positions and to be a target of negative judgment, institutions can help model courage by refusing to cave in to pressure from online agitators, regardless of how vitriolic their words may be. Hardened institutions can cultivate resilience among individuals by normalizing the idea that people ought to be prepared to stand by their ideas even when doing so involves considerable unpleasantness. Even when one is targeted with strawman arguments, associations of guilt by association, and ad hominem attacks, the noble response is to continue seeking truth rather than to falsify one's views or to respond to one's opponents in a similarly fallacious manner.

v.v: The Relative Costs and Benefits of Available Strategies

The previous chapter argued that while banning social media, regulating social media, and voluntarily exiting social media are strategies that have the potential to mitigate the problem of online intimidation culture, they also carry significant costs that ought not be ignored. After having explored the topic of building intuitional resilience to online toxicity, it is reasonable to conclude that this approach to dealing with intimidation culture is far less costly than the alternatives that have been analyzed. The project of hardening institutions does not limit expressive freedom and it does not involve ceding greater power to government offices, which of course might later be abused. Moreover, this project does not involve forgoing the many positive features of social media participation that reasonable people may wish to experience. While hardening institutions will of course require expenditure of time, money, and effort so that policies and protocols can be formulated and implemented in an effective manner, this is a small price to pay when we consider the potential gains that are at stake. If a broad array of institutions throughout society make a clear

commitment to confronting and resisting the pressures of online zealotry, this can have immense benefits for their own internal operations as well as their influence across society more broadly. Institutions can help cultivate an atmosphere of free expression without ever directly altering the conduct of social media companies, and without invoking the power of the state in order to force these companies to comply with particular prescriptions.

All of this is very attractive from a Millian perspective. It was noted in the previous chapter that the value of experimentation and innovation is a key theme that runs throughout *On Liberty*. One of the virtues of addressing intimidation culture via the strategy of building institutional resilience is that it provides plenty of leeway for organizations to identify strategies for coping with online intimidation that are highly efficacious, and are most appropriate for their own needs. It is entirely possible that policies embraced by epistemic institutions such as those that inhabit the realms of academia, journalism, and publishing will need to be different from the policies embraced by businesses that simply wish to offer products and services that are enticing to the broadest possible base of consumers. There is no reason why this should be viewed as a problem from a Millian perspective. So long as institutions are putting measures in place that have the ability to remedy chilling effects and conformity, and invite good-faith dialogue between people with diverging views, then it is not necessary for Millian liberals to take issue with different institutions adopting distinct approaches.

Moreover, since our information ecosystem is constantly evolving, it will likely be necessary for efforts to harden institutions to evolve as well, and continue to respond to new threats and challenges that are not currently visible to those who care about intellectual diversity and

productive discourse. The brightest liberal thinkers of the nineteenth century could not provide direct guidance with respect to social media since they had no familiarity with this form of technology, and likewise, even the most thoughtful and well-intentioned institutional policies that are enacted in the present will likely fail to anticipate novel forms of communication that will emerge in the future, and the challenges that will be associated with them. It is necessary, then, for those who are committed to cultivating resilient institutions to perpetually remain open to revising and updating their preferred policies, while remaining committed to key principles that can help to bring about a substantive atmosphere of free expression.

I have argued that the project of cultivating resilient institutions can produce immense benefits with respect to coping with the phenomenon of online intimidation culture. Importantly, this strategy is worthy of adoption because it avoids the pitfalls associated with alternative strategies that were highlighted in Chapter 4. The chapter that follows will continue to develop this case in favour of the project of constructing and maintaining robust institutions that are sturdy enough to resist various kinds of social pressure. I will endeavour to spell out in greater detail why this project can be of tremendous value to society if it is executed properly. A key goal of this discussion will be to offer an argument regarding the importance of heterodox institutions that will resonate among audiences who may not be as fervent about protecting free expression as Mill and the many philosophers who have been influenced by his works. It will be argued that the project of cultivating heterodoxy among institutions is crucial for the credibility of these institutions, which in turn is integral for the achievement of social trust. Institutions that are vulnerable to social pressure, and prove themselves to be prone to shutting down dissent, are much less likely to enjoy

credibility and social trust than institutions that firmly establish a culture wherein intellectual diversity is welcomed and disagreements are addressed via discussion rather than punishment.

Chapter vi: Heterodox Institutions, Credibility, and Trust

vi.i: Institutional Resilience and Societal Gains

The preceding chapter has argued in favour of embracing a strategy of institutional reform in response to the phenomenon of online intimidation culture. It has been argued that in addition to being efficacious with respect to reining in the pernicious influence of social media, the project of cultivating resilient institutions has the ability to avoid the many pitfalls associated with alternative strategies such as banning social media, regulating social media via the state, and spearheading a voluntary exodus from social media platforms. The present chapter will deepen and elaborate upon this account by drawing attention to the downstream effects of constructing institutions that are heterodox in nature, and are hospitable towards a broad array of ideas and arguments that challenge people in many different ideological camps. More specifically, it will be argued that this program can bolster the credibility of institutions, and assist them in earning the trust of the general public. Since social trust is integral to the cohesiveness and long-term sustainability of society, there are very strong reasons to prevent institutions from becoming conformist monocultures and making a concerted effort to ensure that they remain open to a wide range of arguments and ideas. While the project of increasing institutional credibility and building social trust certainly transcends discussions about social media and its impact on public discourse, my contention is that hardening institutions, and making them more resilient in the face of online backlash, is one key component of this larger project that ought to be appreciated.

If members of the public can observe shared institutions and clearly see that these organizations are able and willing to stand their ground when partisans, ideologues, and bad-faith actors attempt to pressure them into submission using the power of new media, then this will almost certainly bolster people's confidence in these institutions. When institutions are entrusted

with apprehending truths and circulating them throughout society, resistance to external pressure campaigns helps to solidify the notion that the institutions in question are committed to an overarching project that is larger and more important to them than the social currency that might be gained by complying with whichever orthodoxy happens to be dominant at a particular moment in history. An aim of this chapter is to demonstrate that such institutional fortitude can generate increased trust, which can be an enormous gain for society, as an absence of confidence in institutions can entail significant dysfunction and turbulence that is difficult to reverse. The loss of trust in key institutions can have destructive implications for many facets of society, and is nothing to take lightly.

This is especially true when it comes to epistemic institutions, which are responsible for the generation and dissemination of knowledge. These entities provide a proverbial bedrock for productive public discourse by enabling members of society to enter into dialogue with one another about matters of shared interest and concern. The 1967 Kalven Report, which was discussed in Chapter 5, provides a helpful description of what epistemic institutions entail. It has the following to say about universities, which are key institutions of this sort:

> The mission of the university is the discovery, improvement, and dissemination of knowledge. Its domain of inquiry and scrutiny includes all aspects and all values of society. A university faithful to its mission will provide enduring challenges to social values, policies, practices, and institutions. By design and by effect, it is the institution which creates discontent with the existing social arrangements and proposes new ones. In brief, a good university, like Socrates, will be upsetting.[152]

This statement clearly underscores the importance of members of intellectual communities having the freedom to question and scrutinize a vast array of ideas, including ones that are highly cherished. My contention is that such an atmosphere of freedom is integral to the ability of

[152] See Banout and Ginsburg 2024, 164-165.

epistemic institutions to earn the trust of large and diverse populations, and to preserve this trust over time.

While free inquiry is crucial to the proper functioning of epistemic institutions, it is not without qualifications. Epistemic institutions of course ought to adhere to relevant laws, and must refrain from endangering or injuring people for the sake of advancing knowledge. The pursuit of knowledge must be checked by other valuable objectives, such as the preservation of peace and security. Indeed, an epistemic institution that fails to design and enforce basic ground rules with respect to the conduct of its members will almost certainly become dysfunctional, which will undermine its ability to advance knowledge. Sensible rules ought to be conceptualized as tools that can assist the production and dissemination of knowledge, rather than fetters that impede intellectual progress. On this note, it is worth pointing out that the intellectual diversity that is championed by the Kalven committee, Mill, and others, can justify some constraints on the conduct of members of epistemic communities with respect to the ways in which they treat others. Even though members of epistemic communities must have ample leeway in order for these communities to stimulate intellectual growth effectively, some limitations on their freedom are legitimate.

For example: if a member of an epistemic institution engages in conduct that is blatantly bigoted and discriminatory towards others, then this can provide legitimate grounds for disciplinary action or the loss of professional privileges. This is because the behaviour in question is injurious to the integrity of the entire enterprise to which they are ostensibly committed. Participants in intellectual communities are entrusted with the generation and evaluation of ideas,

and this project is undermined when people exhibit prejudice that prevents the ideas of particular persons and groups from receiving a fair hearing. Certain forms of bias and discrimination send clear signals that an individual is not prepared to engage in intellectual inquiry in a serious manner, and therefore can legitimately face repercussions from the intellectual community with which they are affiliated. Again, it makes good sense to view these constraints on people's conduct as assets to intellectual diversity rather than impediments towards it, as they are likely to broaden the range of ideas that are presented and analysed within a particular epistemic community. While intellectual and expressive freedom are key to the pursuit of knowledge, they are not absolutes that trump all other considerations. A balancing between liberty and responsibility is always necessary, and it is incumbent upon epistemic institutions and their personnel to engage in self-criticism in order to determine whether an appropriate balance has been achieved within a particular context.

A key theme of this chapter is that an atmosphere of free expression is crucial to the cultivation and maintenance of trust in institutions. In order to appreciate the importance of this relationship, it is worth taking a moment to consider some potentially troubling questions about the societal costs associated with diminishing social trust in epistemic institutions. What will a society look like if and when vast swaths of its population become distrustful, or even resentful, towards institutions that are relied upon for developing a clear understanding of reality? If epistemic institutions become viewed with suspicion by many ordinary individuals, then how will large and diverse populations be able to communicate and cooperate with one another in a productive manner? It seems that in order for effective deliberation and coordination to take place, a certain amount of trust and mutual understanding between citizens is necessary.[153] As Redstone

[153] Leticia Bode and Emily K. Vraga state: "We know that people tend to be embedded in different information environments, where they may be exposed to more or less misinformation and correction. Likewise, people depend

succinctly states in her book *The Certainty Trap: Why We Need to Question Ourselves More—and How We Can Judge Others Less*: "Societies need a baseline level of trust, in one another and in institutions, to function. That trust underpins a sense of shared goals and a belief that people are mostly working toward a common good. In a democracy, social trust also allows us to live with people with whom we disagree." (2024, 31)

Redstone's discussion is helpful for understanding how pressures from individuals and groups that are highly ideological can derail epistemic institutions and damage the trust in them that is so important. She states: "The erosion of trust comes, at least in part, from having policies on paper that promote free inquiry and expression, making decisions that undermine them, and failing to acknowledge the tension." (2024, 238) She goes on to explain:

> One question that comes up repeatedly when people think about how institutions navigate heated issues is what the institution's goal is. For example, a commitment to understanding what's true about the world might come into conflict with a goal of advancing social progress or creating social change. In this context, many institutions in … higher education, and journalism have, for years, declared truth as the priority. And yet, all of these institutions have found themselves awash in certainty on various heated issues. They have become places where knowledge that should be thought of as provisional is treated as final. (238-239)

The language of certainty is significant here. It can help us understand how epistemic institutions ought to grapple with pressures of an ideological nature. While there is nothing wrong with members of epistemic communities having strong convictions about a variety of topics, there is something seriously wrong with such individuals exalting their own views, treating them as conclusive, and using this attitude of certainty as grounds for shutting down entire areas of inquiry.

on different trusted sources—so a correction from what one group deems a highly credible expert may be seen as a biased or untrustworthy agent by another group. As one example, think about party affiliation in the US: Republicans tend to report lower levels of trust in science and scientists, health experts, and the news media." (2025, 87)

There is a difference between an individual who is a passionate advocate for a cause and an individual who is a dangerous ideologue, and when people become intolerant towards doubt and uncertainty with respect to their own views, they move closer to placing themselves in the latter category. While this critique may sound somewhat harsh, it is firmly in line with the Millian insights and principles that animate this work. Let us consider the following quotation from Mill:

> …it is not the feeling sure of a doctrine (be it what it may) which I call an assumption of infallibility. It is the undertaking to decide that question for others, without allowing them to hear what can be said on the contrary side. And I denounce and reprobate this pretension not the less, if put forth on the side of my most solemn convictions. However positive any one's persuasion may be, not only of the falsity but of the pernicious consequences—not only of the pernicious consequences, but (to adopt expressions which I altogether condemn) the immorality and impiety of an opinion; yet if, in pursuance of that private judgement, though backed by the public judgement of his country or his contemporaries, he prevents the opinion from being heard in its defence, he assumes infallibility. And so far from the assumption being less objectionable or less dangerous because the opinion is called immoral or impious, this is the case of all others in which it is most fatal. These are exactly the occasions on which the men of one generation commit those dreadful mistakes, which excite the astonishment and horror of posterity. (2015, 25)

Mill's warning about individuals who wish to shut down debate due to their own feelings of certainty is helpful for understanding how contemporary epistemic institutions can deal with activists and ideologues. While there is no reason why such persons should automatically be barred from participation in epistemic institutions, there are very strong grounds for putting policies and protocols in place that can prevent such persons from seizing control of institutions and using them as vehicles for their preferred agenda. Once again, Redstone is on point when she states that while she does not have an issue with people being vocal about their beliefs and convictions, she does have an issue with people using a university as a "platform to take a stand on an issue that's clearly controversial, heated, and contentious." (2024, 231) If people attempt to exploit the trust and respect that is placed in epistemic institutions for the sake of promoting their own politics and worldview, they are likely to eventually find this trust and respect being depleted. There are many

ways in which epistemic institutions can protect themselves from becoming dominated by the pressures of activists and ideologues, and a handful of these strategies were outlined in Chapter 2. These strategies include robust protections for employees with respect to their speech and expression, formal opportunities for dialogue and conflict resolution that do not involve punishment, and the adoption of adversarial collaboration projects wherein people are explicitly tasked with working alongside others in order to identify the strongest arguments that can be presented on opposite sides of a debate.

All of these strategies function in order to prevent certain ideas and perspectives from being purged from organizations via punitive action. While it is legitimate for some institutions to exert strict control over the expressive acts of their membership, this is not true of epistemic institutions. The Kalven Report is helpful for appreciating the distinction between epistemic institutions, and other kinds of organizations that embrace a posture that is more activist and ideological in nature. The report states: "A university, if it is to be true to its faith in intellectual inquiry, must embrace, be hospitable to, and encourage the widest diversity of views within its own community. It is a community but only for the limited, albeit great, purposes of teaching and research. It is not a club, it is not a trade association, it is not a lobby." (164-165) This statement provides strong support for the notion that epistemic institutions can be strengthened by promoting an atmosphere wherein dissent and intellectual diversity are vigorously protected. If an institution that is entrusted with generating and disseminating knowledge degenerates into a "club", "association", or "lobby", then it will become disconnected from its proper mission and will become an epistemic institution in name only. The pursuit of knowledge is not compatible with the doctrinaire promotion of a particular belief or worldview. This is one reason why documents such as the Kalven Report can

be viewed as valuable tools for hardening epistemic institutions and preventing them from becoming hijacked by activist and ideological pressures, and thereby losing the trust of laypeople.

This line of reasoning is of course highly congruent with Mill's social and political philosophy, which has informed earlier chapters and helped us arrive at a clear understanding of what is at stake in debates about online intimidation culture. A Millian atmosphere of free expression cannot be cultivated and sustained when epistemic institutions are malfunctioning, and different segments of a population become unable to reach agreement about fundamental empirical matters. While Mill is indeed an advocate for intellectual diversity, it would be erroneous to think that his arguments in support of free and open discourse imply that a fractured and disorderly epistemic landscape is in any way desirable. Instead, it is far more plausible to think that a chaotic information environment is a barrier to the productive discourse between diverse individuals and groups that Mill wishes to promote. My view is that online intimidation culture is a major factor that has played a role in the severe loss of trust in epistemic institutions that has unfolded in recent history, and one objective of this chapter is to demonstrate how and why this is the case.[154] Accordingly, it is incumbent upon those who wish to address the decline of social trust in institutions to confront and address social media's corrosive impact on these entities and the discourse that surrounds them.

[154] According to FIRE scholar Sarah McLaughlin, universities are aware of the reputational damage that can be inflicted through social media: "Brand protection efforts extend to social media … A 2020 survey of over 200 universities found that three in ten schools employ a private, curated blacklist to block comments on their Facebook pages. It is no surprise that these tools were often used to hide commentary that could be harmful to their reputation." (2025, 100)

vi.ii: Eroding Trust in Institutions

It is well documented that public trust in a wide array of institutions has declined sharply over the course of recent decades in liberal democracies. Philosopher Kevin Vallier begins his book *Trust in a Polarized Age* with the following observations:

> Americans are finding it harder and harder to trust one another. Social trust in the United States, our trust in our fellow citizens, has fallen dramatically. In the early 1970s, around half of Americans said that most people can be trusted. Today that figure is less than a third. Political trust, trust in government and democracy, has fallen steeply as well. Throughout the 1960s, over 70 percent of Americans said they trusted government in Washington always or most of the time. By the early 1990s, that number had fallen below 30 percent, and after a brief rebound in the early 2000s, it has collapsed to 17 percent as of 2019. More troublesome still is that Americans reporting no confidence at all in their national government doubled from around 14 percent between 1995 and 2011 to 28 percent in 2017. We see a similarly disturbing pattern in partisan distrust. In 2017, around 70 percent of Republicans said they distrusted anyone who voted for Hillary Clinton for president; likewise, around 70 percent of Democrats said they distrusted people who voted for Donald Trump. People not only distrust politicians from other parties, they distrust anyone who votes for the other party, that is, many millions of people. In our politically polarized age, we trust each other less simply based on how we vote. (2020, 1)

This data about the decline of trust in governmental and representative institutions, as well as fellow citizens, is congruent with the concerns about polarization that have animated this discussion of social media and online intimidation culture. It is clear that the rise of social media has been accompanied by a simultaneous rise in social alienation and antagonism that is too significant to ignore.[155] While there is little doubt that multiple factors have contributed to this alienation and antagonism, it is appropriate to explore the question of whether the decidedly toxic dynamics of social media discourse have played a role in producing this state of affairs. Vallier is

[155] Alex Edmans notes the link between polarized public discourse and social media: "Public discourse is increasingly polarized, with opinions formed on ideology, not evidence. The most pressing issues of our time, such as climate change, inequality and global health, are steeped in falsehoods. In the past, we knew what the reliable sources were, such as a doctor or medical textbook for health advice and an encyclopaedia for general knowledge. Now one half of Americans obtain news 'often' or 'sometimes' from social media, where false stories spread further, faster and deeper than the truth because they're more attention-grabbing." (2024, 9-10)

far from alone in viewing these trends as cause for alarm; scholars in other disciplines have similarly raised concerns about the ubiquitous loss of trust that is unfolding throughout democratic societies. The economist Benjamin Ho expresses worry about this matter in his aptly-titled book *Why Trust Matters: An Economist's Guide to the Ties that Bind Us*. He notes that while from a broad, macro perspective, people have become immensely more trustful of one another as the evolution of human cultures has progressed, this expansion of trust has recently been challenged. He states that "…despite this millennia-long pattern of expanding trust, there has been a recent hiccup in this trend. Even as our trust in each other has grown, our trust in experts and institutions has begun to falter. This erosion in trust of politicians, scientists, doctors, and economists has been well documented in recent decades" (2021, 183). While Ho highlights the fact that "the development of technology expanded our circle of trust from the confines of our immediate family and tribe to more and more of the global community", he also provides warnings about "the role of technological innovation in undermining trust", which align with the overarching argument of this discussion.

It makes good sense for Ho to draw attention to the loss of trust that has taken place with respect to experts and institutions outside of government, as these are entities that are frequently called upon to make consequential decisions that can shape the lives of many millions of people. The fact that experts and institutions are losing the trust of the general public is a concerning development that raises deep questions about the sustainability of the large-scale cooperation that is required in modern societies.[156] We have noted that if members of a society lack trust in a broad

[156] Ben-Porath offers commentary about the importance of trust for democracy: "Trust in fellow citizens, as well as trust in our power to hold institutions accountable to our needs and interests, is key to the functioning and sustainability of democracy. A major way to respond to concerns about the erosion of trust, as noted by Robert Putnam and many others, is the threading of the civic fabric through local ties and interpersonal as well as institutional connections …

range of core institutions, at a certain point it will become difficult for society to function even at a basic level. A foundation of trust is required in order for people to coexist and cooperate in large groups, and current trends raise the possibility of such coexistence and cooperation deteriorating. It is not hyperbole to state that the erosion of trust in core institutions particularly undermines the long-term survival prospects of liberal and democratic societies, as these societies are by definition comprised of large collections of ordinary individuals coming together to participate in governance alongside their peers. Lynch provides a view of democratic politics that is instructive for our purposes:

> I … use the term 'democratic politics' to mean inclusive, representative, and respectful deliberation between free and equal persons about what society ought to do in the face of collective problems. Democratic politics in this sense is not associated with any particular political party. It is a kind of practice, or way of interacting politically, that can take place in or out of formal democratic arrangements. But it is the kind of politics practiced whenever democratic societies are spaces of reasons, aspiring to support a kind of public sphere—a space where disagreements can be navigated with reasons as opposed to violence or manipulation. (2025, 14)

This process of deliberation and shared governance is liable to break down when people become unable to agree about which sources of information ought to be viewed as credible, and begin to view people with different worldviews as threats to their own security and wellbeing. There is no doubt that many different kinds of institutions must enjoy a certain level of trust in order for social unrest and dysfunction to be avoided. That being said, it is worthwhile to consider which *kinds* of institutions are most vital to the fundamental operations of liberal polities. While the loss of trust in governmental and representative institutions is certainly worthy of attention and concern, one can plausibly argue that epistemic institutions play an even more fundamental role in

Extreme partisanship, which drives the political sphere to focus on mere power and undermines the feasibility of collaboration (or its perceived electoral incentives), is detrimental to generating trust. More alarmingly, this process can lead to polarization and mistrust around facts and about the institutions and standards that are used to help sort out core facets of a shared epistemology." (2023, 14-15)

the organization of society. This is because these are the institutions that people look to in order to develop a basic understanding of the world around them and guide their decision-making. As philosopher Jason Stanley states: "Democracy requires a common shared reality, including a common understanding of the past …Without such an understanding, one cedes power to hierarchy, or potentially an autocrat." (2024, 183-184)

While an inability to agree about matters of public policy can be troubling, there are even stronger grounds for alarm when members of a society lose their ability to agree on basic facts, as this can set the stage for intense conflict and instability. Indeed, this is one reason why the issues of misinformation, disinformation, and conspiracism have inspired a vast academic literature over the course of the 2010s and 2020s. DiResta appropriately laments the social and cultural fragmentation that has flourished in the age of digital media:

> Shared reality has splintered into bespoke realities, shaped by recommendation engines that bring communities together, filled with content curated from the media and influencers that the community trusts. Very little bridges these divides. This has profound implications for solving collective problems or reaching the kind of consensus on which democracies depend. (2024, 11)

While the arguments about institutional credibility and trust that are presented in this chapter are relevant to a wide range of institutions, epistemic institutions are the ones that are brought to the fore due to their critical role in laying the groundwork for communication, deliberation, and cooperation between citizens of various polities.

vi.iii: Epistemic Institutions and Online Intimidation

Let us return again to the issue of online intimidation culture. Having noted the dramatic decrease in trust that has unfolded in recent decades, it is reasonable to ask whether the dynamics

of online discourse have collided with epistemic institutions and social trust in destructive ways. It seems that this is very likely the case.[157] It is not too difficult to see why the social media dynamics that have been highlighted and criticized throughout the previous chapters could inflict damage on the credibility of epistemic institutions and undermine their ability to maintain the trust of the general public. We have seen that social media discourse incentivizes maximalism, tribalism, and antagonism, and that the vitriolic character of online communication can play a role in establishing an atmosphere of intimidation wherein people avoid exploring challenging questions and ideas out of fear of being targeted with castigation and exclusion.[158] Unfortunately, this pattern of behaviour is the precise opposite of what is needed in order for epistemic institutions to demonstrate their credibility and earn the trust of large and diverse populations.[159]

Some of the most prominent and esteemed epistemic institutions in the world are academic institutions. These entities are entrusted with producing original research, engaging in rigorous peer review, edifying pupils, and broadening society's wealth of knowledge by exploring new and challenging intellectual horizons. While these institutions are some of the most respected and influential that societies have to offer, they are not immune from the decline in trust that has been

[157] Haidt explains: "Recent academic studies suggest that social media is indeed corrosive to trust in governments, news media, and people and institutions in general … The literature is complex—some studies show benefits, particularly in less developed democracies—but the review found that, on balance, social media amplifies political polarization; foments populism, especially right-wing populism; and is associated with the spread of misinformation." (2022)

[158] David Zweig notes the human tendency to conform to group pressures: "… most (though not all) humans do not well tolerate being seen as contrarian to their group, or even thinking contrary to their group. This is obvious through all of history, with a rich literature on in-group-/out-group bias, social identity theory, peer pressure, and groupthink among peoples in nations, religions, and professions." (160-161)

[159] Lukianoff and Schlott argue that cancel culture is a key factor undermining trust in academia: "As meaningful debate putters out and Cancel Culture thrives on campuses, the rest of the country has taken notice. Increasingly, Americans are distrustful of academia as an institution … Cancel Culture has devastated the trust we have in the very institution we rely on to produce knowledge… and to educate future generations of Americans." (2023, 62)

taking place over recent decades. Former Harvard University president Derek Bok, a lawyer, describes this phenomenon:

> In addition to the threat of government regulation, public opinion has been shifting in ways that make elite universities more vulnerable to political intervention. While higher education enjoyed a favorable reputation in America for many decades, trust in universities began to slip early in this century. This trend has accelerated in recent years. The percentage of Americans who believe that colleges and universities have a positive effect 'on the way things are going' in this country dropped from 69 percent in 2020 to 58 percent in 2021 and again to 55 percent in 2022. Meanwhile, the gap between the levels of trust in universities from Democrats and Republicans has widened in recent years to such a point that well over half of all Republicans now lack confidence in higher education. Several factors appear to contribute to the erosion of public trust. The constant rise of tuitions at a faster rate than increases in the cost of living has doubtless been a contributing factor, as has a growing concern that colleges may not be preparing students adequately for employment. The exceptional loss of trust among many Republicans stems in part from the vast predominance of liberals in college faculties and the belief that professors 'bring their political views into the classroom.' (2024, 196-197)

Bok helpfully points to a handful of factors that have likely played a role in the general public losing its confidence in the ability of top universities to have a positive impact on society. This downward trajectory is striking in light of the fact that universities have historically been viewed as institutions that operate above and away from the kinds of parochial concerns that often shape the domains of commerce and electoral politics, among others. While universities imposing unreasonably large financial burdens on students is certainly a serious issue, the issue of growing politicization is more relevant to our concerns in this discussion. It is clear from Bok's comments that a large portion of the general population perceives the realm of higher education as being increasingly and excessively politicized, meaning that ideas and perspectives that are worthy of consideration are being dismissed for the sake of protecting established orthodoxies within academic disciplines.

It is interesting for our purposes that elsewhere in his discussion, Bok specifically highlights the role of social media in generating an atmosphere of intimidation and self-censorship wherein community members refrain from engaging sincerely with one another and addressing contentious issues. Bok states:

> …many students self-censor, because they are reluctant to express unpopular views that may provoke disapproval by their classmates on a variety of sensitive subjects … Instructors cannot prevent such reactions, but they can at least encourage students with unpopular views to speak up and engage in a discussion. … There are organizations outside the university that unleash torrents of abuse through social media on faculty members or students who express whatever these groups regard as biased or offensive views. Such tactics are unfortunate. Nevertheless, while campus officials can issue statements reassuring aggrieved students and faculty that they disapprove of shaming or threatening messages, universities have no power to punish such behavior or remove the harm they cause. (2024, 123)

The relationship between social media and campus life has been well-documented by scholars in various disciplines, and it is the core theme of Redstone and Villasenor's book *Unassailable Ideas: How Unwritten Rules and Social Media Shape Discourse in American Higher Education.* These authors take a firm stance regarding the power of online communications to influence educational institutions:

> …social media can act both directly (e.g., through call-out campaigns) and indirectly (through behavior modification aimed at avoiding social media opprobrium) to shape what happens on campus. One reason is that technology has upended how everyone—including the academic researchers we entrust to discover and disseminate new knowledge and the professors and other instructors we entrust to teach college classes—communicates. (2020, 43)

I obviously share this view, and think that it provides a valuable avenue for exploring the issue of social trust in academic institutions and how it can be restored and promoted. Here we can clearly see how the dynamics of social media have the potential to undermine trust in academic

institutions.[160] If social media toxicity has the potential to constrain academic discourse through intimidation, and members of the public are losing trust in academia precisely because they perceive its intellectual culture as being excessively narrow and rigid, then it is reasonable to conclude that social media is playing a role in the loss of social trust that universities are experiencing.

While Bok states that there is little that universities can do to address the attacks that take place via social media, the analysis presented in the previous chapter, as well as the present chapter, offers a very different view. My alternative position is that hardening institutions and cultivating a culture of openness and heterodoxy can go a long way in reducing the power of social media attacks, even if such strategies cannot eliminate such attacks entirely. Institutions do not need to shut down or reshape social media discourse in order to remedy the pervasive chilling effects generated by online vitriol. Rather, they can combat online intimidation culture by taking proactive measures to establish an atmosphere of free expression wherein people with many different views are welcome to participate in academic activities without fear of formal or informal punishment. A key goal of this chapter is to explicate the societal benefits associated with the construction and protection of epistemic institutions that are decidedly heterodox in nature, and to make a case in favour of incorporating Millian principles and precepts into the routine operations of epistemic institutions so that they can successfully generate social trust, thereby preventing societal dysfunction of various kinds.

[160] Lukianoff and Schlott note: "If we want a better society that produces better solutions to the problems it faces, we need to be teaching nonconformity at every single level of the education process … And yet our education system is incentivizing conformity and groupthink. Unless this environment drastically improves—and quickly—we shouldn't be surprised that trust in the accuracy of professors' and experts' findings diminishes. Mistakes abound when groupthink goes unchallenged." (2023, 78-79)

Academic institutions are important in myriad ways, but other kinds of epistemic institutions deserve consideration as well. We can glean insight about online intimidation culture by examining the operations of journalistic institutions in order to appreciate how the pressures of social media can undermine processes that are conducive to the maintenance of social trust. It goes without saying that journalistic institutions are key epistemic institutions. They are relied upon by many millions of people in order to develop an understanding of current events in their home countries and abroad. Journalistic institutions have enormous influence in the domains of electoral politics, business, international relations, and so on. It was noted in Chapter 3 that over the course of the 2010s and 2020s, journalists have often faced pressure to adhere to the prevailing orthodoxies of their social milieu, and that social media has played a role in fuelling these pressures. Research indicates that the world of journalism has become thoroughly intertwined with the world of social media, and that nearly every facet of the journalistic enterprise has been influenced by social media in some way.

Rose-Stockwell provides a vivid portrait of social media's corrosive influence on the journalism profession. He explains: "In the United States, nearly every journalist uses social media. Editors use it every day to decide how to allocate their coverage of critical issues. The industry of journalism has been consumed by the social media news feed." (2023, 49-50) He goes on to highlight how even journalists have succumbed to the addictive features of social media discourse: "An economic dependency has taken hold, as those who are responsible for sourcing truth have become professionally and personally addicted to these tools. Many journalists even see tweets as equally newsworthy as headlines from the Associated Press. This vastly increases the risk of bad

ideas, fringe content, and false news becoming amplified." The author continues by noting: "This is a painful open secret in the news business because today the hidden governors of our information system have become algorithms. They're built by humans to capture attention at almost any cost. And increasingly, that cost is our civility, decency, and measured discourse." (2023, 50)

While it would be reasonable to view these dynamics as good reasons for journalists and the institutions with which they are affiliated to distance themselves from social media platforms, it is unfortunately the case that individuals involved in this industry have generally been unable, or unwilling, to abstain from participation in social media discourse despite the many drawbacks that it entails. It is plain to see why social media platforms are attractive for people who are involved in the journalism profession. During the 2010s, Twitter in particular became a staple of journalists' professional routine, as they could use this platform to keep up with current events, communicate with sources, and market their work to broad audiences in order to increase its reach and impact. When used skillfully, social media platforms can benefit the careers of news media personnel, and assist consumers of news in finding informative content that is most relevant to their needs and interests. Accordingly, it would be an error to assert that social media platforms have had an entirely negative effect on the domain of journalism. Unfortunately, it is also the case that social media platforms have provided a space for communication wherein journalists can face enormous peer pressure.

Journalists who broach ideas and subjects that are frowned upon by other members of their profession can swiftly be castigated for falling afoul of entrenched social norms. While on the surface it might appear as though social media platforms provide a venue for freewheeling debate

about a vast collection of subjects, the enormous conformist pressures that permeate these platforms have made them into a space wherein influential users with large followings can effectively police the views of others, assembling groups of social media users to lambast them as a form of social punishment if and when they fall out of step with dominant norms and expectations. Journalists are generally well aware that other journalists are active on social media, and it can take considerable courage to resist the pressures of groupthink and be willing to highlight facts and arguments that have the potential to upset popular narratives. While ordinary people face enormous pressure to conform thanks to social media, journalists arguably face even greater pressure due to the fact that they are immersed in networks of relatively opinionated and articulate persons who can use their significant platforms to broadcast their views to large audiences. The risk of reputational damage and ostracism is high for people who are active in such networks.

Moreover, it is worth noting that social media platforms have introduced incentives into the realm of journalism that encourage professionals in this area to be increasingly blatant when it comes to broadcasting their ideological and partisan allegiances.[161] While it was once at least somewhat taboo for journalists to make their personal political views known to news audiences, the dynamics of social media can generate significant rewards for journalists who take on a clearly ideological and partisan posture. After all, we have seen that one of the best ways to achieve salience and approval on social media is to launch attacks on others who are viewed as unsympathetic targets. Since the realm of social media is ripe with competition for attention and status, journalists who wish to establish a presence on social media are motivated to demonstrate

[161] Taibbi states: "In the age before social media, most reporters didn't have to expose their political opinions to the world. Today everyone is effectively an op-ed writer. [Liz] Spayd's take was, this isn't necessarily a good idea, and exposes both reporters and papers like the *Times* to accusations of bias in ways we never had to worry about before." (2021, 90)

to their peers and their audiences that they deserve the respect and acceptance of their respective ideological networks on the grounds that they hold appropriate views and are loyal to the right causes.

In order to understand how social media platforms have the potential to exacerbate partisan and ideological bias in journalistic institutions and their personnel, we can consider the work of Ho. This researcher notes the significance of "overconfidence" with respect to political matters. He states:

> Research shows that overconfidence in political beliefs translates into more extreme ideological points of view. In a Pew Research center survey in which Americans were asked to rate how warmly they felt about members of the other political party on a scale from 0 to 100, the warmth they felt toward the other party declined from close to 50 in the 1970s to under 30 in 2016. (2021, 210)

He explains how the surge of digital media that has unfolded over recent decades has made people much more likely to dismiss ideas that challenge their existing beliefs:

> Social media exposes us to more and more stories designed to confirm our existing beliefs that make us more and more confident. If we are more confident about what we believe, then we are less likely to trust someone who tells us something that contradicts our beliefs. That increases the incentive to conform. If saying what is seen as the wrong thing gets you labeled as someone not to be trusted for having bad values, then we have extra incentive to only say the 'right thing.' (2021, 212)

While partisan journalism has undoubtedly existed for a very long time, and no journalistic outlet has ever been perfectly truthful and free from bias, it is plausible to argue that the rise of social media and its enormous influence over the journalistic profession has encouraged journalistic institutions to become increasingly removed from the goals of neutrality, fairness, and objectivity that at one time were considered paramount for the journalistic enterprise. These institutions and their personnel want to remain in the good graces of people that will respond with

irritation if they are presented with information and arguments that call their worldview into question. As many laypeople have been encouraged to espouse increasingly zealous positions, so too have journalistic institutions and the personnel that operate within them. Mounk offers the following comment about how social media has shaped journalistic outlets:

> The story of how the dynamics of social media transformed public discourse starts, at the beginning of the decade, with the rise in prominence of seemingly niche platforms like Tumblr and Thought Catalog. It culminates, at the end of the decade, in seismic changes at the most influential media outlets, from the BBC to NPR and MSNBC.

He explains:

> Anyone who compares a copy of *The New York Times* or *The Guardian* in 2010 with a copy of those same newspapers in 2020 would be struck by the difference in their tone and content. One small indication of this transformation lies in some of the articles and op-eds that would have been considered too extreme to see the light of day a decade earlier. (2023, 93)

Additionally, DiResta offers the following sober analysis about the media coverage that surrounded a video of a tense interaction between a high school student and an elderly Indigenous man that received enormous attention via social media:

> … for multiple days, many thousands of people on the internet, hundreds of influencers, and then numerous media articles dissected this extremely small moment of tension, something that, I would argue, never needed to be an online moment at all. Nothing had really happened. No one was injured. No one had to fixate on these people or go dig through their lives to find out who they really were. Three groups of people had a short disagreement, a few moments of real-life tension. What resulted was a pseudo-event, a spectacle, something that influencers called attention to and media covered in ways framed to appeal to their particular audiences. Depending on whom you followed, you heard about it at a different time, saw it described in very different ways, and heard that the 'other side' was a bunch of liars, manipulators, and villains. Depending on which influencers and outlets you trusted, you formed a particular view of events, who was right or wrong in the situation, and what their online punishment should be. Social media's curation algorithms, particularly trending topics, push these pseudo-events… into the fields of those who have previously engaged with similar types of content—bait dangled at those mostly likely to take it. While I would argue that this mess was bad for everyone involved, it was great for platform engagement and influencer engagement, and the online crowds got some excitement out of a morning of righteous indignation. (2024, 75)

Given what we know about the dynamics of online discourse, it is reasonable to suggest that social media, and the intimidation culture that thrives on social media, are likely important factors in the decline of trust in journalism that has taken place in recent decades. As journalistic institutions and their personnel have become more comfortable with courting the approval of social media audiences, the public has taken notice of the increasingly ideological character of news coverage that purports to be nonpartisan. Political scientist Jacob Hale Russell and legal philosopher Dennis Patterson offer the following powerful critique of the manner in which the journalistic profession has evolved, and specifically raise the idea that "overreach" among elites has led to public distrust:

> The fear of bothsidesism has begun to undermine journalism, an institution that traditionally prided itself on doing precisely that—giving a voice to 'both' sides. Eliciting contrary viewpoints from sources doesn't require that journalists repeat unchecked falsehoods or falsely imply that all sides share equal popularity. But that is not enough to satisfy younger journalists … In surveys, they openly aver that equal coverage of divergent political views isn't warranted. The main effect of this stance is, more than anything else, to confirm populists' sense that elites don't want to hear from them. … Skepticism is mislabeled denialism; dissent is censored as misinformation or derided as conspiracy thinking; open discussion is marked off bounds as bothsidesism. Expertise is replaced with a pale simulacrum that consists more of name-calling, tone policing, and censoring than of engaging in dialogue. In other words, elites are building even higher ramparts around their epistemic edifice to protect the very overreach that caused the public to distrust them in the first place. (2025, 198)

It would of course be reductive and wrongheaded to suggest that online intimidation culture is singlehandedly responsible for this troubling decline in trust.[162] However, when one considers the trends that have unfolded in recent decades throughout the field of journalism and the incentives that have driven them,[163] it becomes quite clear that social media dynamics have steered

[162] Even if we set aside concerns about social media, it is still evident that journalistic institutions often produce coverage that will be flattering towards the audiences that sustain these institutions financially. Ho states: "The logic is simple: people like reading news that confirms their prior beliefs. Therefore, pandering to those beliefs sells more newspapers." (2021, 205)

[163] Caitlin Petre states: "… large technology platform companies, with their enormous numbers of highly engaged users and the ability to target them with unprecedented precision, have proved irresistible to the very advertisers who previously relied on news organizations to disseminate their messages. Large technology platforms such as Facebook

journalists away from the important project of building their credibility by consistently delivering news coverage that is as accurate and fair to diverse stakeholders as possible, and towards the alternative goal of building brands for themselves that will enable them to win prestige and advance their careers. Social media provides journalists with a virtually endless supply of feedback towards their reporting, opinions, and online posts, and unfortunately, listening to this feedback has the potential to derail the pursuit of truth and replace it with the pursuit of popularity, affirmation, and financial rewards.

The upshot is that the dynamics of social media that reward tribalism and conformity have the ability to shape the conduct of important epistemic institutions, just as they have the ability to shape the behaviour of ordinary individuals going about their daily lives. While epistemic institutions and their personnel may enjoy considerable esteem in many respects, they are not immune from the pressures and incentives that often cause laypersons to behave in ways that are detrimental to themselves and others. The pressures of online intimidation culture have the ability to erode trust in epistemic institutions by flattening them into monocultures wherein intellectual diversity and rigorous debate are increasingly deemphasized, and personnel are instead encouraged to pursue social rewards and professional advancement by aligning themselves with prevailing orthodoxies, and dismissing information and arguments that might present a challenge to these orthodoxies. While the above commentary has focused on the domains of journalism and academia, similar arguments could be made about a broader range of epistemic institutions.

and Google have … become crucial intermediaries between news and audiences, rendering media companies dependent on their mysterious algorithms for online distribution." (Petre 2021, 25)

vi.iv: The Vulnerability of Experts

The relationship between epistemic institutions and social media is a contentious subject. A skeptic might assert that journalistic institutions and their personnel are exceptionally vulnerable to the conformist pressures that pervade social media, and that people involved in different epistemic institutions, such as those committed to scientific research and its publicization, operate in a very different manner. One might make the case that expert individuals, and organizations that are involved in intensive empirical research, cannot be swayed as easily as those who are involved in the dissemination of news content. Unfortunately, researchers have noted that experts, including professionals involved in the scientific enterprise, have not been immune to the social media dynamics that have been discussed throughout this and previous chapters. Despite their considerable intelligence, experts are also vulnerable to online intimidation culture. They too can succumb to the peer pressure and groupthink that pervade social media discourse, potentially disrupting the pursuit of truth and accuracy.

In his book *Within Reason: A Liberal Public Health for an Illiberal T*ime, epidemiologist Sandro Galea argues that experts who challenge prevailing ideological commitments can quickly be punished via social media, further underscoring this technology's power as an engine of conformity:

> Peer review is a means of testing our scientific conclusions to ascertain their integrity and support better scholarship. In recent years, forms of media have begun to take the place of peer review in shaping the trajectory of our thoughts. Peer review continues, of course, but far more influential in some ways are the feedback loops enforced by media bubbles and social media platforms like Twitter, where public health practitioners are rewarded for expressing ideas that fall within certain ideological parameters and punished for straying outside them. Where peer review helps sharpen our pursuit of truth, the media often amplifies distorted or incomplete thinking, undermining the intellectual foundations of our field. (2023, 20)

The notion that academic peer review is informally being replaced by social media feedback is a bold and disquieting one.[164] This is not something that anyone who cares about epistemic institutions and the integrity of public discourse ought to take lightly. If it is true that even highly credentialed people like professional academics can fall victim to the bad incentives that permeate social media, then the ability of these platforms to inflict damage on an array of influential institutions throughout society must be viewed as a real problem. Insofar as we rely on these institutions for achieving a shared sense of reality that can facilitate communication and cooperation, the damage inflicted upon these institutions by the forces of intimidation culture can properly be understood as a threat to societal stability and flourishing. Our institutions will not be able to live up to the lofty and noble objectives that they are officially committed to if their personnel are more concerned about winning the approval of online and offline crowds than producing quality work that can assist these institutions in realizing their overarching missions.

Later on in his discussion, Galea explicitly highlights the importance of social trust in science, and how it can be undermined by inappropriate influence from the emotional content that pervades social media:

> … it is particularly important for scientists to resist the influence of emotion and social media's tendency to strengthen it. Because science is supported by an empirical framework with a rich history of guiding human inquiry, there is an assumption that scientific conclusions are less subject to the influences that shape a tweet or a newspaper editorial. When these influences do start to shape scientific discourse, and when this influence becomes clear to the wider public, it can be corrosive both to scientific output and to the trust this output has historically engendered. This can erode the foundation of data necessary for making informed, rational decisions about health within a liberal framework. For this reason, when scientists engage on social media, we should take care that we

[164] Zweig raises concerns about healthcare professionals being subjected to social punishment for questioning orthodox views: "If you worked at a large hospital, there were severe professional repercussions for speaking against the CDC, or the views of your colleagues, your bosses, or 'the narrative.' … Multiple experts I interviewed had been reprimanded by superiors for publicly questioning, either on social media or via interviews with the press, the restrictions on children and school closures, or for pointing out inconvenient data." (2025, 159)

> amplify the best of scientific rationality rather than ideology and emotion. In a time when technology has given everyone a voice, it is up to us to use ours to help advance the data-informed clarity that allows us to make the best possible decisions about health. (2023, 71)

It is evident that this author has a robust understanding of the incentives that animate social media discourse, and also understands the idea, articulated in Chapter 3 of this work, that the corrosive influence of social media can easily spill over from the online domain to the offline domain. This analysis from Galea coheres with that offered by the scholars Stephen Macedo and Frances Lee of Princeton University. They argue that public discourse effectively broke down during the COVID-19 pandemic, as many views that deserved a serious hearing were dismissed by powerful individuals and groups, including experts, and cast as immoral. (2025, 21-22) Most importantly for our purposes, these authors argue that public discourse during this period needed to be much more hospitable to the views and concerns of laypersons, who were unjustifiably excluded from complex and consequential policy debates. While Macedo and Lee do not use the language of "intimidation culture" in their discussion, it is clear that they too are concerned with the atmosphere of fear that was established during the pandemic era and its ability to shut down constructive dialogue. Interestingly, these authors specifically cite Mill in their discussion of how powerful individuals and organizations abused the power with which they were entrusted during this period. Let us consider the following passage:

> Remarkably quickly, in late spring and summer 2020, the consensus that emerged among … policymakers and opinion leaders hardened into the dismissal of dissent. It was insisted that there was only one way forward and that those who argued for greater attention to the costs of business and school closures or the efficacy of masks were guilty of callous indifference to human life. This insistence was false and deeply unfair: human lives and well-being were at stake on both sides of these policy debates, as should have been recognized at the time. University researchers who dissented from Blue-state orthodoxy were subject to vilification and even faculty censure. The premature moralization of disagreement along partisan lines undermined basic norms of democracy (accountability requires an opposition party), liberalism (openness to criticism as the best test of truth, per John Stuart Mill), and science (a community of scientists open to a diversity of viewpoints,

> contestation, and refutation, per Karl Popper and Oreskes). Elite conventional wisdom needed to be checked by discussion among a broader range of experts, and—perhaps even more important—it needed to be balanced by a broader discussion extending beyond knowledge workers and professionals to include citizens more generally. (2025, 21-22)

These comments from Macedo and Lee provide a helpful illustration of how intimidation culture, which is an abstract concept, can have very real implications with respect to how societies grapple with concrete issues, including global emergencies. Moreover, their message about the importance of including ordinary citizens in discussions about contentious issues is highly congruent with the normative philosophical views that animate this discussion. Public discourse simply cannot function adequately when vast populations are afraid of voicing their ideas and asking challenging questions out of fear of punishment. While experts are capable of succumbing to the pressures of intimation culture, they are also capable of perpetuating intimidation throughout society and effectively excluding ordinary people from public discourse. In societies wherein very many people feel that epistemic institutions and their personnel are intolerant of scrutiny from outsiders, it should come as no surprise when social trust deteriorates. It would be a profound error to overlook the role that an atmosphere of intimidation has played in undermining trust over the course of recent decades. My hope is that we can now see that while our media ecosystem is not singlehandedly to blame for this loss of trust, the perverse incentives that it introduces into public discourse can and do influence institutions and their personnel in damaging ways.

vi.v: Intimidation Culture and Institutional Credibility

In light of these considerations, it is appropriate to shift our focus towards the positive project of building the credibility of institutions so that trust in them can be cultivated and maintained over long periods of time. Rather than continuing to catalogue the many ways in which

online intimidation culture has corroded epistemic institutions and undermined the public's trust in them, it is more productive to explore the question of how institutions might go about reversing this alarming trend. My view is that the social goods associated with free expression that were explicated in earlier chapters can provide a helpful guide to cultivating institutions that have the ability to shore up their credibility over time and win the trust of members of the public. While there is no doubt that many bad actors can and do make deliberate efforts to hurt epistemic institutions and their relationship with the general public, and that this merits serious analysis, it is nonetheless true that epistemic institutions can do a great deal in order to minimize the impact of such bad actors. If epistemic institutions make a concerted effort to build resilience to the forces of intimidation, bolster their credibility, and generate social trust, then the detractors who wish to cause them harm can be disempowered, as fewer members of the public will be receptive to messaging that seeks to stoke backlash towards said institutions. My hope is that in addition to resonating with individuals who are passionate about free expression and open inquiry, the arguments offered here will pique the interest of researchers and commentators who are concerned about salient issues such as misinformation, disinformation, conspiracism, and the rise of the "post-truth" era. My contention is that the cultivation of institutional credibility is among the most powerful antidotes available for grappling with these issues.

It was argued in Chapter 1 that engagement can be understood as the currency of social media platforms. In order to develop and deepen our account of how epistemic institutions can go about shoring up public confidence, it may be helpful to conceptualize credibility as the currency of institutions. Practically by definition, credibility is what is required in order for institutions to maintain the trust of diverse collections of people over years, decades, and centuries. When

institutions clearly articulate their objectives, remain committed to these objectives over long periods of time, and convey the results of their efforts to onlookers, they augment their credibility. This necessitates a certain amount of humility: when institutions accept accountability for errors and shortcomings, without offering excuses or engaging in spin, this too bolsters their credibility by demonstrating to the public that they are capable of engaging in self-criticism and reform when these are necessary.[165]

While credibility is a concept that is impossible to quantify precisely, this concept is valuable because it prevents us from becoming excessively fixated on external factors that can shape social trust in institutions. Some might argue that when laypeople lose trust in epistemic institutions, this effectively signals an increase in cynicism or ignorance among the general population that ought to be addressed. Indeed, there may be merit to these kinds of diagnoses with respect to the issue of declining social trust. Perhaps it really is the case that this withering of trust is fuelled by changes in the attitudes of ordinary people that ought to be criticized. However, if we solely view the issue of declining social trust from a lens that draws attention to the shortcomings of people who exist outside of epistemic institutions, then we run the risk of overlooking the ways in which these institutions might be able to reverse this decline by revising their own conduct. Invoking the concept of credibility in debates about epistemic institutions and their relationship with the general public is worthwhile because it can prevent researchers from developing a myopia

[165] Russell and Patterson highlight the importance of humility among those who lay claim to expertise: "Expert overreach deploys a veneer of expertise to marginalize or censor dissenting views. This distortion of expertise claims certainty rather than facing up to complicated unknowns, and it blames all failures on those who ignore the supposed expert consensus. It removes tough questions and difficult, even tragic trade-offs from their proper sphere of political judgment and public debate. Such overreach cannot be said to be the genuine exercise of expertise because to truly use knowledge requires perspective, integrity, and humility." (2025, 40)

that leads them to exaggerate the role of laypeople, and downplay the role of institutions and their personnel, when it comes to the issue of decaying social trust.

As we have noted, epistemic institutions are entrusted with the generation and dissemination of knowledge. They are supposed to be organizations that many different kinds of people can rely upon in order to provide accurate information that can aid them in their own thinking and decision-making. Even if we accept the notion that no institution can be perfectly neutral, it is nonetheless reasonable for people to expect epistemic institutions to prioritize accuracy above all else, and to be willing to produce and share information that is potentially controversial or uncomfortable to contend with. When epistemic institutions channel their time, energy, and financial resources towards the goal of producing high-quality information that is verifiable by others, then they are acting in accordance with their normative role. This means that epistemic institutions are emphatically not supposed to take on a didactic role wherein they use their vast resources and influence in order to propagandize the general public. Moreover, it is not for epistemic institutions to pick and choose which information it is appropriate for the public to hear, on the grounds that certain kinds of information might lead people to embrace the wrong ideas and behaviours. When epistemic institutions lose touch with their appropriate role in society[166] and begin to engage in these kinds of efforts to steer society towards particular ideas and conclusions that they have deemed to be of importance, then they increase the likelihood that their credibility will be eroded.

[166] It was noted above that the mission of epistemic institutions provides justification for constraints on the behaviour of participants. Biased and discriminatory conduct was considered in particular. Similarly, it is reasonable to point out that the mission of epistemic institutions can provide grounds for them to make efforts to recruit members of marginalized groups that are underrepresented in their operations. Attempts to increase the demographic diversity of the institution's personnel are consistent with intellectual diversity so long as the institution continues to remain neutral on contentious matters, and permits its members to engage in freewheeling debate.

Macedo and Lee provide commentary that underscores the importance of separating the objectives of epistemic institutions from other kinds of objectives related to promoting particular kinds of social and political causes:

> Some evidence suggests that today's scientists are more inclined than those of the past to censor research they perceive as socially harmful. In November 2023, several dozen scholars of social psychology and other social sciences penned a joint article in the Proceedings of the National Academy of Sciences (PNAS), showing that 'scientific censorship is often driven by … pro-social concerns for the wellbeing of human social groups,' among other motives. Scientists and other experts may seek to block the dissemination of ideas by both 'hard' and 'soft' means, with the latter including 'social punishments' such as ostracism and shaming. Moralized bias manifests itself in other ways as well, including in judging and assessing evidence: '96% of statistical errors directionally supported scientists' hypotheses, suggesting credulity among scholars toward favorable outcomes.' (2025, 283)

These statements make it clear how the phenomenon of intimidation culture intersects with the credibility of epistemic institutions. If it is indeed true, as evidence suggests, that the personnel of epistemic institutions are willing to use "ostracism and shaming" as tools for censoring research that they view as problematic, then it is plain to see how intimidation can stifle inquiry and steer members of intellectual communities away from their appropriate objectives. If shaming and ostracism are undermining the ability of epistemic institutions to produce and disseminate accurate research, then this is likely to damage the credibility of these institutions. This is why it is sound to reach the conclusion that intimidation culture is one factor, among others, that has the potential to erode institutional credibility.[167] The reasoning here is that when institutions and their personnel compromise the project of generating and disseminating knowledge in order to appease actors who

[167] Diana C. Mutz notes: "Incivility among political advocates…produces systematically less trust in government than equivalent disagreements that transpire more politely. Clearly, there is something about incivility that rubs Americans the wrong way. Not only attitudes toward politicians and Congress, but also levels of support for the institutions of government themselves were influenced." (2015, 89)

threaten them with social punishment, this sends a powerful signal throughout society that these institutions cannot be trusted when the going gets tough, so to speak. It may be the case the institutions fare adequately in low-pressure environments wherein criticism and pushback are minimal, but when these institutions must face difficult decisions about how to navigate significant backlash from confrontational critics, their inability to stand firm can speak volumes about their normative role in society and the extent to which this has been compromised.[168]

In light of these comments about how the dynamics of intimidation culture can severely damage the credibility of institutions, it is appropriate to take this opportunity to consider how the project of hardening institutions and building their credibility is connected to the three social goods associated with freedom of expression that were enumerated and analyzed in Chapter 2 of this work. My contention is that appreciating these social goods can play an important role in developing a clearer understanding of how institutional credibility can be strengthened in practice. Rather than simply flagging the importance of institutional credibility and its relationship to the modern phenomenon of online intimidation culture, I wish to offer a more specified account of how the Millian philosophy of free expression that animates this work can provide a guide to reforming institutions in beneficial ways, and reversing the decline in trust that has taken place

[168] It must be acknowledged that epistemology is a vast branch of philosophy, and plenty of disagreement exists about generating and disseminating knowledge. Advocates for standpoint epistemology, to take one example, may take issue with the portrait of epistemic institutions that has been offered in this chapter on the grounds that it does not adequately grapple with philosophical concepts such as social situatedness and epistemic privilege. While I cannot answer these challenges in detail here, my view is that liberal approaches to epistemology are capable of accommodating concerns regarding the necessity of giving a fair hearing to the voices of the marginalized and ensuring that opportunities for uptake are available to them. Rauch's commentary is instructive: "We learn empirically that women are as intelligent and capable as men; this knowledge strengthens the moral claims of gender equality. We learn from social experience that laws permitting religious pluralism make societies more governable; this knowledge strengthens the moral claims of religious liberty. We learn from critical argumentation that the notion that some races are fit to be enslaved by others is impossible to defend without recourse to hypocrisy and mendacity; this knowledge strengthens the moral claims of inherent human dignity. Over decades and centuries, ethical concepts about gender equality and religious liberty and individual dignity emerge, evolve, and stand the test of time. They are not empirical knowledge, to be sure, but they are subject to social checking; as a result, they are knowledge, and they exhibit progress." (2014, 173)

throughout recent history.[169] The spotlight here will continue to be placed on epistemic institutions, although these arguments may be relevant to other kinds of institutions as well to varying degrees.

Let us begin with the social good of critical intellectual faculties. In order for credibility to be maximized, it is important for people outside of institutions to see that the personnel operating within them are not only able and willing to hone their critical intellectual faculties, but also that the deployment of these faculties is welcomed by the broader culture and milieu of the relevant institution. We have seen that the development of critical intellectual faculties necessitates the questioning of many rival ideas, including ones that are popularly held and viewed as paramount. This is a core insight that drives Mill's advocacy for freedom of thought and expression. If people are barred, officially or unofficially, from probing ideas and following their curiosity,[170] then their critical thinking abilities will be cramped as a consequence. If the general public perceives that the personnel of epistemic institutions are not encouraged to deploy and refine their critical intellectual faculties, and instead are rewarded simply for parroting fashionable ideas and talking points, then this will damage the credibility of these institutions. Credible epistemic institutions ought to welcome and nurture individuals with sharp intellects who wish to ask difficult questions and challenge dogmas.

[169] It is instructive to consider Mill's optimistic view about what can be achieved when people are willing to examine competing ideas: "there is always hope when people are forced to listen to both sides; it is when they attend only to one that errors harden into prejudices, and truth itself ceases to have the effect of truth … And since there are few mental attributes more rare than that judicial faculty which can sit in intelligent judgment between two sides of a question, of which only one is represented by an advocate before it, truth has no chance but in proportion as every side of it, every opinion which embodies any fraction of the truth, not only finds advocates, but is so advocated as to be listened to." (2015, 51)

[170] Just as ordinary expressive acts ought to be constrained by Mill's harm principle, the same is true of intellectual pursuits as they are described here.

Importantly, the public should be able to see not only that members of epistemic communities are in possession of strong critical intellectual faculties, but also that these faculties are being deployed towards the appropriate overarching objective: apprehending truth. If critical intellectual faculties are deployed in a perverse fashion that is disconnected from the pursuit of truth, then this too will have the potential to seriously undermine institutional credibility. If the personnel of epistemic institutions are in possession of sharp and powerful intellects, but deploy them mainly for inappropriate purposes, such as advancing their own position within the relevant institution and reaping rewards, then this too will damage institutional credibility and social trust. Indeed, members of the public might be even more suspicious of highly intelligent people than others who are less gifted, as people in the former category are likely to be better at deploying language in manipulative ways and constructing rationalizations for bad behaviour. It is a major issue when the personnel of epistemic institutions and communities lose touch with the basic overarching project of generating knowledge and disseminating it throughout society. While it is certainly not the case that every single layperson who observes epistemic institutions has a detailed understanding of their inner workings, it is reasonable to think that when epistemic institutions become disconnected from their core mission, this loss of focus and purpose will be detected by a significant portion of the general public, thereby undermining trust.

The above comments about the honing and development of critical intellectual faculties are closely linked with the social good of authenticity in discourse. Authenticity is enormously important in order for epistemic institutions to be viewed as credible. If the public can see that members of epistemic communities are engaged in the promotion of ideas not because they sincerely believe them to be accurate, but rather because they are aligned with a particular ideology

or orthodoxy, then trust will be severely undermined. As Mill observes in his writings, people are generally highly attuned to social norms regarding which words and ideas are acceptable and which are grounds for social punishment. This means that an atmosphere of intolerance within institutions can subvert authenticity in discourse, and cause personnel to engage in preference falsification, misrepresenting their own views.[171] The deployment of social punishment as a means of narrowing debate and inquiry is highly likely to make the public lose trust in the proclamations made by epistemic institutions and their personnel. Laypeople are capable of understanding that when critics and dissidents are marginalized or ejected from intellectual communities, the voices that remain in good standing within these communities will be those of conformists who do not have the desire or the fortitude to challenge dominant perspectives.

The notion that criticism and dissidence ought to be purged from institutions for the sake of bolstering their image in the eyes of the public is wrongheaded, because intolerance of criticism and dissidence is widely understood to be a sign of institutional weakness. Redstone argues compellingly that a culture of certainty can actually damage trust in institutions. She states:

> Certainty…leads to the erosion of trust in institutions by, as with individuals, weaving its way into contentious social and political issues. It leads institutions to take positions on heated issues, without even necessarily realizing that's what they're doing. And when this is done by an institution that the public expects to either be unbiased or to welcome and be open to a wide range of perspectives, trust is eroded. This is the inevitable result of the public seeing that the institution is not living up to the values it claims to hold. (2024, 216)

Institutions that welcome their own personnel, as well as laypeople who are not directly affiliated with them, to voice their questions and concerns openly and without fear will do a much better job of building up their credibility than those that meet any significant scrutiny with

[171] See Timur Kuran's *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (1998).

dismissiveness or hostility. If epistemic institutions wish to enjoy esteem in the view of the public and preserve their positions of authority over the long term, one of the last things that they should do is send out a powerful signal indicating that they are incapable of coping with challenging questions and rival perspectives. If they are interested in building their credibility rather than corroding it, then epistemic institutions ought to welcome authenticity in discourse rather than attempting to police discourse through formal or informal means.

This brings us to the third social good that has been brought to the fore throughout this dissertation: that of equity in accountability. This social good is also very relevant to discussions about institutional credibility and how it can be enhanced. In order for intellectual communities to be credible, the general public needs to see that community members are called to account on a consistent basis regardless of their worldviews or ideological orientations. Again, arguments presented by Redstone are instructive. This author offers an insightful account of how the process of challenging ideas in an equitable manner can help cultivate trust:

> …when the questioning and challenging of ideas is done with honesty and sincerity—what some people refer to as 'good faith'—we can both transform conversations and build trust. What's more, there can be a sense of balance in knowing that, in a battle of ideas, the process of questioning values, beliefs, and principles applies to everyone equally. (2024, 55-56)

Redstone provides an important reminder that engaging in debate does not mean that one is without convictions: "Being able to name our values and allow them to be challenged while understanding we do not need to let them go is an important skill. It's why a commitment to questioning and challenging our thinking doesn't mean we can't declare right or wrong." (2024, 56) These comments are highly congruent with the Millian principles and precepts that inform this discussion. It is plain to see that shielding certain views from scrutiny, while welcoming

scrutiny towards other views, is a kind of hypocrisy that is likely to severely damage trust in institutions, and especially epistemic institutions. In order for these institutions to have credibility over the long term, they must demonstrate that people of all kinds are welcome to participate in the process of questioning and reason-giving that is characteristic of constructive public discourse.

Equally important is for such institutions to uphold equity in accountability when participants violate the rules and norms that govern them. For example: if people with one ideological orientation are punished in large numbers for relatively minor forms of misconduct, while people with a more popular and orthodox ideological orientation face no repercussions for similar misconduct, then this inequity will undermine the credibility of epistemic institutions. This is because such a pattern of behaviour clearly conveys that a double standard is present within the relevant institution, and that it is indirectly policing speech by subjecting individuals with different views to inequitable treatment. Such inequitable treatment can effectively exclude certain views from consideration within an institution, as the people who hold them will routinely be marginalized, and perhaps even ejected from the relevant community altogether. It is worth noting that this issue is distinct from the related issue of whether institutional responses to various kinds of failures on behalf of personnel ought to lean towards harshness or towards lenience. The point is that once policies and protocols are put in place regarding how individuals ought to conduct themselves, all personnel ought to be held to account in an equitable manner.[172] Indirectly punishing certain members of an intellectual community for their ideological orientation is likely

[172] To appreciate the importance of rules being enforced in an impartial manner, we can look to Vallier: "The chief touchstones of trustworthiness in a diverse political order are its rights practices: observable acts of publicly protecting and exercising basic rights. Free speech is a rights practice where individuals exercise their liberty to speak under protection from coercive interference from others and from government … My argument is that the institutions of an open society, that is, a society with a broad range of liberal rights practices, have the unique power to sustain trust between diverse perspectives, overcome the illusion of culpable dissent, and stop political war." (2020, 22)

to undermine public trust in institutions by signalling that they are willing to exert influence over their intellectual climate through means that are decidedly duplicitous and underhanded. If the public notes that people with unpopular views are held to more stringent standards than those with popular views, this too will damage the credibility of epistemic institutions and their associated intellectual communities.

All three of the social goods that have been brought to the fore in this discussion are integral to the project of developing epistemic institutions that are equipped to earn the trust of a diverse public that is constantly evolving, and to maintain this trust over long periods of time. While it is entirely possible that other social goods can also inform the policies and governance of epistemic institutions, the ones that have been explored in this discussion are particularly helpful with respect to keeping the pressures of conformity at bay and ensuring that institutions function as spaces wherein heterodoxy is welcomed rather than thwarted. My view is that organized, institutional efforts to rein in intimidation and promote intellectual diversity can play a key role in remedying the worrying loss of trust that many researchers have observed throughout recent decades. The social goods of critical intellectual faculties, authenticity in discourse, and equity in accountability can provide a guide as to how institutions can go about strengthening their credibility and maintaining the trust of diverse individuals and groups.

vi.vi: An Antidote to Institutional Capture

I have argued that key social goods associated with free expression can function as desiderata informing the design of institutions in order to remedy the problem of declining trust. At this juncture, I would like to say more about how the establishment of epistemic institutions

that are heterodox in nature can operate as an antidote to the societal dysfunction that takes root when an atmosphere of intimidation, rather than an atmosphere of free expression, pervades society. A key problem associated with intimidation culture and its ability to quell dissent is that it facilitates the capture of epistemic institutions by actors that are not committed to appropriate objectives. While people can debate the finer details about how institutional capture ought to be conceptualized, the core point for our purposes is that this phenomenon involves institutions being steered away from their core objectives and towards alternative ones that are deemed paramount by individuals that wield formal or informal power with respect to the relevant organization. While institutions and their personnel do have a responsibility to provide space for activist messaging to be expressed and heard, they do not have a responsibility to take positions on the innumerable contentious issues that permeate society. Following the prescriptions of the Kalven Report, institutions should function as venues that house speakers, rather than as speakers themselves. Given the arguments presented throughout this chapter, it should now be clear enough that when epistemic institutions become dominated by persons who wish to advance a particular social or political agenda, rather than the generation and dissemination of knowledge, this undermines institutional credibility and the trust that comes with it.

This loss of credibility and trust is certainly an important matter. However, the pernicious impact of institutional capture does not end there. There is another major issue at play in debates about institutional capture that deserves to be highlighted. Specifically, when institutions cave in to activist pressures, effectively allowing themselves to be directed by partisans and ideologues, this incentivizes opposing partisans and ideologues to fight aggressively to take over said

institutions so that they can advance their competing agenda.[173] It is not the case that once an epistemic institution becomes dominated by a specific partisan or activist movement, it can simply continue to operate without fanfare. To the contrary, when influential institutions are captured by ideological interests, this can motivate other societal actors with rival interests to do everything in their power to seize control of these institutions so that they can be the ones in charge of policing the organization's operations, and dictating which questions and ideas are to be tolerated.[174] When institutions become compromised by partisan and ideological objectives, this encourages various factions throughout society to battle for power and influence so that they may fashion these institutions towards their own ends.

If institutions become viewed by the general public as ideological monocultures that are hostile to skeptics and dissidents, then we can expect certain individuals and groups to feel maligned and slighted as a result. People are unlikely to respond well when they perceive that their own ideas and perspectives are being unfairly suppressed rather than being given a fair hearing. This can motivate aggrieved parties to eventually seek retribution if and when they find themselves in a more powerful cultural position. Instead of being shunned and excluded, they can then be the ones doing the shunning and excluding. Institutional capture by one party can encourage and beget institutional capture by opposing parties. It can encourage people with different worldviews to

[173] Grossmann and Hopkins offer insight: "Citizens of both parties dislike the perceived politicization of major institutions even when they see those institutions as aligned with their own beliefs. Americans are less likely to trust social institutions across journalism, science, government, the corporate sector, and nonprofits that they view as having become politicized." (2024, 198-199)

[174] Jacob Mchangama explains that in the United States, political actors have responded to censoriousness with more censoriousness: "… Republican lawmakers supposedly worried about cancel culture's effect on free speech took to fighting fire with fire. In states like Oklahoma and Florida, Republicans have proposed removing critical race theory from classrooms, willfully ignoring that government-mandated restrictions on curricula in and of themselves create the risk of establishing a particular form of ideological orthodoxy." (2025, 340)

jockey for institutional power so that they can defeat their rivals and promote their own orthodoxies. Even if people agree in principle that a particular institution ought not become politicized, when given a choice between being in a position of power or a position of submission, they are likely to choose the former option. Institutional capture accordingly promotes social strife and the turbulence that comes with it. This is deeply concerning. We do not want aggressive pendulum swings in culture shaping the policies of epistemic institutions, especially when large and diverse populations must rely upon them in order to understand and navigate the world.

This is a powerful reason for constructing institutions that are heterodox in nature, and for people with many different worldviews to welcome the presence of heterodox institutions throughout society. Rather than turning institutions into battlegrounds for ideologues, a better option is to construct institutions wherein intellectual diversity is expected and promoted so that many different kinds of voices can receive a fair hearing on a consistent basis. Once again, it is helpful to consider the words of the Kalven Committee:

> The neutrality of the … institution arises … not from a lack of courage nor out of indifference and insensitivity. It arises out of respect for free inquiry and the obligation to cherish a diversity of viewpoints. And this neutrality as an institution has its complement in the fullest freedom for its faculty and students as individuals to participate in political action and social protest. It finds its complement, too, in the obligation of the university to provide a forum for the most searching and candid discussion of public issues.[175]

The institutional neutrality prescribed here decreases the incentive for activists of various kinds to fight for control over institutions so that they can use them as platforms for the promotion of their own worldview. It is arguably already the case that aggressive pendulum swings are taking place with respect to the realm of elite post-secondary education, as politicians have sought to

[175] See Banout and Ginsburg 2024, 165.

address perceived political bias and unfairness by interfering with the autonomy of universities and subverting academic freedom and open inquiry.[176] Harvard University and Columbia University in particular have become flashpoints wherein political actors have sought to respond to censoriousness with a different brand of censoriousness. This indicates that the phenomenon of institutional capture is far from being merely a theoretical possibility. It is incumbent upon those who care about expressive freedom, social trust, and the like to think carefully about how it can be addressed.

vi.vii: An Antidote to Siloing

In addition to examining how heterodox epistemic institutions can help combat institutional capture, it is worthwhile to consider how such entities can help to alleviate siloing,[177] which is an issue that has been mentioned in previous chapters. Siloing takes place when individuals and groups with different worldviews become disconnected from one another, and inhabit intellectual communities that are increasingly insular. Vallier explains how a decline in trust can fuel the process of siloing:

> Increases in divergence … may decrease trust because people may be less likely to trust persons who have values distant from their own. As cultures develop social markers for identifying these diverse groups, and as content creators draw our attention to those markers by representing or exaggerating them in news, TV, movies, and social media, we will tend to culturally sort ourselves into different social silos that seldom interact with one another. (2020, 8)

[176] Bok explains: "The … disadvantage from the near absence of conservative faculty is the risk that Republican politicians will intervene to try to compensate for the political imbalance among university professors. Governor Ron DeSantis of Florida (a graduate of Yale College and Harvard Law School) has led the way in this endeavor." (2024, 90)

[177] Lynch alludes to the issue of siloing: "We are used to living in our bubbles of hyperpartisan information, protected from the distasteful opinions of those different from us. And we've become numb to the seemingly endless stream of conspiracy theories and outright falsehoods sloshing around the internet…" (2025, 34)

We can see that the issues of social trust, media dynamics, and siloing are all interconnected. Plenty of evidence indicates that social siloing is a real phenomenon that is shaping culture and politics, and that it is facilitated by modern channels of communication. Author Tom Nichols has the following to say about the matter:

> A 2021 study led by researchers at Princeton found that the self-sorting process is now practically a reflex, with people now sorting themselves into silos or 'epistemic bubbles' online without even realizing it. Worse, multiple studies find that access to broadband connections actually *increases* political polarization, because—unsurprisingly—people go online to find others who share their preexisting views, and thus to confirm and strengthen their biases rather than to interact with those who disagree. This unwillingness to hear out others not only makes us all more unpleasant with each other in general, but also makes us less able to think, to argue persuasively, and to accept correction when we're wrong. When we are incapable of sustaining a chain of reasoning past a few mouse clicks, we cannot tolerate even the smallest challenge to our beliefs or ideas. This is dangerous because it both undermines the role of knowledge and expertise in a modern society and corrodes the basic ability of people to get along with each other in a democracy. (2024, 135)

This analysis from Nichols is highly relevant to the arguments advanced in this chapter. The dynamics of our contemporary information environment are causing people with different ideologies and worldviews to increasingly avoid interaction with one another and seek out spaces wherein they can maximize interaction with those they view as allies. These patterns of social organization fuel tribal behaviour wherein people who think differently from one's self are perceived not as interlocutors who have the ability to help sharpen and refine one's own worldview, but as antagonists who are potentially dangerous and must be defeated. It is clear that siloing is a phenomenon that has the ability to seriously undermine a culture of free expression wherein people seek out a broad array of perspectives in the interest of expanding their intellectual horizons and achieving a better understanding of the truth.

My contention is that the project of hardening institutions, and making them more resistant to various kinds of pressure and backlash, can help rein in the intellectual siloing that Vallier, Nichols, and others correctly worry about. If institutions make unambiguous commitments to protecting their personnel from attacks, including ones involving online mobs, this will signal to observers that people can function inside of these institutions without necessarily conforming to every idea that happens to be dominant at a particular moment in time. Rather than seeking safety in numbers and aligning themselves with organizations wherein their own ideas are dominant, individuals will be encouraged to operate inside of institutions that are more diverse in nature, thereby curbing siloing and its pernicious effects.

It is worth pointing out that the construction of new, alternative epistemic institutions is not necessarily an effective remedy for the problem of siloing. People who are serious about cultivating an atmosphere of free expression and its associated social goods should not want segments of society with different ideological orientations to build opposing epistemic institutions that refuse to interact with one another. This is a form of division and balkanization that can fuel the division that so many researchers and commentors wish to remedy. Political scientists Matt Grossmann and David A. Hopkins suggest that the establishment of rival epistemic institutions with different ideological orientations can fuel polarization:

> Conservatives … responded to the liberal dominance of the university system by building think tanks in Washington … as well as in state capitals around the nation. Think tanks developed in emulation of universities – complete with resident scholars, named fellows, and resources to develop expert knowledge … Political scientist E. J. Fagan has investigated the historical development of partisan think tanks and held them partially responsible for increased polarization in Congress. (2024, 199-200)

The impulse to construct new institutions that can compete with flawed existing institutions is certainly understandable, but the risk of facilitating siloing through such measures should not be overlooked.

While there will always be a place in liberal societies for publications and organizations that are explicitly ideological in nature, such as *Reason* (a libertarian outlet), *Jacobin* (a socialist outlet), and *National Review* (a conservative outlet), there must also be a place for institutions that are hospitable and accessible to a broad array of perspectives, wherein people with different ideas can collaborate in order to identify truths. This is a point that was raised in the discussion of supporting institutions in Chapter 2, and it is worth exploring further here. As we have seen, while it is important to have diversity between institutions, it is no less important to have diversity within institutions, especially when they are explicitly tasked with the generation and dissemination of knowledge.[178] Troublingly, one can plausibly argue that the phenomenon of siloing has already manifested itself in the social media sector. Since Twitter was acquired for 44 billion dollars in 2022, its usership has changed considerably. Let us recall that in response to this acquisition by the world's wealthiest individual, many former Twitter users have made the decision to migrate to other social media platforms such as Bluesky, Mastodon, and Threads. We now have a situation in the social media sector wherein Twitter (known as "X" since 2023) is a platform with a palpable right-wing inclination, and Bluesky in particular has emerged as a rival platform with a palpable left-wing inclination. For those who view the issue of siloing as some kind of remote theoretical

[178] Lee McIntyre similarly notes: "More speech across diverse outlets does not balance out disinformation, because if no individual network has to be “fair,” this incentivizes news siloes that are devoted to skewed content, which is sometimes all that anyone watches. As we have learned in the last decade, when it comes to factual information (and not just opinion- or editorial-based content), balance *across* media sources is not nearly as effective at preventing disinformation as balance *within* media sources." (2023, 76)

possibility, this reality ought to be sobering. It indicates that it is indeed very possible for intellectual communities to split into competing networks that are actively hostile towards one another, and can even encourage one another to become more extreme.

The issue of extremity has been a theme of previous chapters, and it is no less relevant in discussions about heterodox institutions and the dynamics of siloing. A very good reason to cultivate heterodox institutions and prevent siloing of the media ecosystem is the problem of group polarization. This is a well-studied and well-documented phenomenon that has manifested itself in a number of countries. Sunstein explains that when people are given opportunities to interact with others who resemble themselves in terms of their opinions and beliefs, this can cause them to ultimately embrace a more extreme and uncompromising outlook:

> What happens within deliberating bodies? Do groups compromise? Do they move toward the middle of the tendencies of their individual members? The answer is now clear, and it is perhaps not what intuition would suggest: members of a deliberating group typically end up in a more extreme position in line with their tendencies before deliberation began. This is the phenomenon known as group polarization. Group polarization is the usual pattern with deliberating groups, having been found in hundreds of studies … a group of people who think immigration is a serious problem will, after discussion, think that immigration is a horribly serious problem; that those who dislike the Affordable Care Act will think, after discussion, that the Affordable Care Act is truly awful; that those who approve of an ongoing war effort will, as a result of discussion, become still more enthusiastic about that effort; that people who dislike a nation's leaders will dislike those leaders quite intensely after talking with one another; and that people who disapprove of the United States, and are suspicious of its intentions, will increase their disapproval and suspicion if they exchange points of view. (2021a, 79-80)

This information ought to be alarming for those who care about the integrity of public discourse. If siloing of the media ecosystem continues to take place, we can expect many of the issues that have been highlighted in this dissertation to become even more intense. Partisans will become increasingly intolerant and distrustful towards their ideological adversaries, and they may

even lose the ability to debate one another if they begin to adopt conceptual frameworks and vocabulary that are alien to competing camps. If we fail to cultivate heterodox institutions wherein a vast array of ideas can be studied and challenged, then this will set the stage for a period of escalating tension and dysfunction as individuals and groups with different worldviews become siloed into increasingly insular intellectual communities wherein members are pressured into embracing increasingly extreme – and in many cases, unreasonable – ideas and behaviours.

Accordingly, perhaps it is unsurprising that in addition to being an expert on the phenomenon of group polarization, Sunstein is an advocate for free expression and heterodox institutions. He argues forcefully that institutions ought to encourage participants to express themselves in a candid and sincere manner, instead of permitting social pressures to generate conformity and extremity:

> It is extremely important to devise institutions that promote disclosure of private views and private information. Institutions that instead reward conformity are prone to failure; institutions are far more likely to prosper if they create a norm of openness and dissent. The point very much bears on the risks of group polarization. Groups of like-minded people are likely to go to extremes, simply because of limited argument pools and reputational considerations. The danger is that the resulting movements in opinion will be unjustified. It is extremely important to create 'circuit breakers' and to devise institutional arrangements that will serve to counteract movements that could not be supported if people had a wider range of information. (2021a, 148)

This argument from Sunstein is highly relevant to the issue of intimidation culture that lies at the heart of this discussion. It is obviously the case that institutions that allow a culture of intimidation to thrive will incentivize their members to conform, and even to misrepresent their own views for the sake of evading punishment. This can fuel group polarization and societal dysfunction more broadly. In Chapter 2, I raised the idea that instead of expelling people who voice ideas that are deemed offensive or wrongheaded, societies should strive to foster their

inclusion so that these people can be exposed to others who may act as beneficial intellectual influences. Hopefully this idea seems more plausible now that we have examined the phenomenon of group polarization and considered its relationship to the subject of institutional design. If we construct institutions that easily cave in to social pressure, including social pressure that operates via social media, then we will create an environment in which people who are expelled or silenced by institutions are increasingly motivated to construct their own intellectual communities wherein they can express themselves without fearing backlash or retribution. This siloing will produce a set of serious societal harms. The cultivation of heterodox institutions can help societies avoid this unfortunate outcome, and protect the integrity of public discourse while shoring up social trust.

In line with Sunstein, Russell and Patterson offer a clear account of what can go wrong when debate becomes stifled by dogmatism. They argue that when groups of people are driven to extremes due to the absence of scrutiny and dissent, this can lead to the implementation of flawed policies that are difficult to reverse:

> When we drive out uncertainty and debate and falsely or prematurely declare consensus or that a question is 'settled,' we make it more likely that a mistaken policy will be widely adopted in its most extreme form. Policy is sticky, and bad policy can be hard to undo. We also make it far less likely that research will be done to evaluate whether a given policy decision was correct. And the public is misled about the true basis for policy decisions, which ultimately rest not just on neutral facts but on the political preferences of those who anoint themselves the keepers of the facts. (2025, 40)

Accordingly, the dual project of hardening institutions and cultivating heterodox institutions ought to be fuelled by an understanding that when institutions degenerate into monocultures, this can lead to errors that will take years or even decades to properly address. When epistemic communities fail to deploy the mechanism of open inquiry and debate, which is a powerful vehicle for self-correction, they are likely to produce serious error and the dysfunction that comes with it.

Therefore, it is in our interest to design institutions that are informed by a Millian philosophy of free expression, and are capable of maintaining a heterodox character, even in the face of significant blowback and criticism.

Chapter vii: The Promise of Social Media

vii.i The Shape of Social Media to Come

It goes without saying that the bulk of this discussion has been highly critical of social media platforms, the incentives that they introduce into public discourse, and the tactics that people deploy in order to ascend the social media ranks in pursuit of engagement and financial rewards. Accordingly, it seems appropriate to conclude this discussion of free expression in the age of social media by exploring the positive potential of this technology with respect to the enhancement of expressive freedom and the facilitation of meaningful interaction between diverse individuals and groups across the globe. While it is true that social media has had a corrosive impact on public discourse in many ways, at no point in this discussion has it been denied that this technology also has many benefits. Perhaps more importantly, it has been acknowledged that there is little reason to think humans have finished the project of creating and reforming social media institutions. The social media sector remains dynamic and unpredictable in many ways. Since people are capable of learning from their mistakes, as Mill and many others emphatically point out, it is entirely possible that the social media industry will strive to deliver services to consumers that are more prosocial in character than the ones that we have critiqued over the course of the preceding chapters. If consumers become sufficiently fatigued and frustrated with the cacophony that pervades the social media status quo, they may very well be eager to experiment with platforms that steer users towards informative and fulfilling online experiences rather than performative quarrelling.

Our interest in this chapter is not in detailing the countless ways in which social media services might be designed or revised in order to deliver a better overall experience to users and to society more broadly. This is a subject of discussion that could fill volumes. Since online

intimidation culture is a product of the incentives that are present in online discourse, realigning these incentives in thoughtful ways could go a long way towards producing a social media ecosystem that is less combative than the one that we have become accustomed to. Decentralizing content moderation could help, as it would give ordinary users more control in establishing and maintaining online communities that revolve around specific topics and interests. Eliminating features such as "trends", which amplify incendiary content and encourage social media users to participate in mob behaviour, could also play a role in reducing the toxicity of online communications. Perhaps most obvious of all, de-boosting inflammatory content, thereby reducing its reach, is another salient strategy that has the potential to change the tone and tenor of social media discourse. The upshot is that reforming the social media landscape is something that will require trial and error in addition to creativity and good intentions. It is fortunate that many people inside and outside of the social media industry understand that there are deep problems with the manner in which these platforms function, and are interested in cultivating healthier dynamics in online spaces.[179]

While all of these ideas, and many more, are worthy of serious attention and debate, the main goal of this chapter is to consider the societal gains that might be achieved through the process of connecting people around the world via the relatively new technology of social media. Instead of simply taking for granted that more connection is better than less connection, as many social media optimists have in the past, the objective here is to provide a philosophical account of

[179] Bail emphasizes the importance of empirical observation with respect to this project: "As we unlock the keys to make our platforms less polarizing, we can use insights from social science to make them a reality. Instead of implementing untested interventions proposed by technology leaders, pundits, or policy makers, we must build the methods of empirical observation of human behavior into the architecture of our platforms, as some social media companies have already begun to do. Along the way, we must recognize that the immense challenges we face will continue to evolve over time." (2021, 131-132)

why and how increased communication between individuals and groups in distant locales could make societies better off. Once again, the philosophy of Mill will help guide and inform this discussion. Specifically, we will examine Mill's views about the importance of consensus-building, as well as his ideas about humans' immense capacity for compassion. It will be argued that social media platforms have the potential to be transformative and beneficial with respect to the project of arriving at robust consensuses between diverse individuals and groups, as well as the project of expanding compassion. The technology of social media can invigorate public discourse within societies and between societies, and it would be an error to let the pernicious phenomenon of online intimidation culture cause us to overlook these potential gains.

vii.ii The Project of Consensus-Building

At this point, it is appropriate to consider an objective that is outlined in Mill's texts, and one that is frequently overlooked. This is the project of consensus-building. We can understand consensus-building as the process of getting people with diverse worldviews and perspectives to arrive at broad agreement about matters that were previously contentious. Due to the fact that Mill is a staunch advocate for intellectual diversity and a culture of openness, it is easy to misconstrue his views about the importance of achieving societal consensus. Some might assume that his philosophy of free expression encourages a state of affairs wherein different individuals and groups espouse opposing ideas about important matters, and use the liberties that they have been afforded in order to communicate about these areas of divergence in perpetuity, thereby avoiding the forging of a consensus. This assumption is mistaken, as textual evidence clearly indicates that Mill views the achievement of societal consensus as something that can be highly desirable, so long as this consensus is forged in an appropriate manner. This is evident in the following passage:

> Is the absence of unanimity an indispensable condition of true knowledge? Is it necessary that some part of mankind should persist in error, to enable any to realize the truth? Does a belief cease to be real and vital as soon as it is generally received – and is a proposition never thoroughly understood and felt unless some doubt of it remains? As soon as mankind have unanimously accepted a truth, does the truth perish within them? The highest aim and best result of improved intelligence, it has hitherto been thought, is to unite mankind more and more in the acknowledgement of all important truths: and does the intelligence only last as long as it has not achieved its object? Do the fruits of conquest perish by the very completeness of the victory? I affirm no such thing. As mankind improve, the number of doctrines which are no longer disputed or doubted will be constantly on the increase: and the well-being of mankind may almost be measured by the number and gravity of the truths which have reached the point of being uncontested. The cessation, on one question after another, of serious controversy, is one of the necessary incidents of the consolidation of opinion; a consolidation as salutary in the case of true opinions, as it is dangerous and noxious when the opinions are erroneous. (2015, 43)

Mill states that as human societies develop and improve over time, they will reach agreement about a greater number of matters, and fewer subjects of discussion will be sites of controversy.[180] This statement about the value of consensus-building might seem curious coming from someone who is so ardent in defending diversity of thought and opinion, but it also makes good sense in light of Mill's confidence about the ability of humans to apprehend truths via the deployment of reason and vigorous debate. The most logical way to read Mill is to conclude that the diversity of thought and opinion that he champions are valuable precisely because they can assist populations in building consensuses around ideas that are sound, and capable of withstanding intense scrutiny. A robust and resilient consensus around an idea cannot be reached until every relevant contrary view has received a fair hearing, and this is why freedom of thought and

[180] Philosopher Michael Fuerstein notes that societal consensus can be reached while still leaving plenty of space for intellectual diversity: "…public opinion polling shows very significant convergence of moral attitudes on the central issues in … historical examples. Support for school desegregation in the U.S. now approaches 100%. Support for same-sex marriage is currently at 71%. Those who approve of a woman working even when 'she has a husband capable of supporting her' was at 82% as of the late 1990s. These are genuine examples of significant, and in the first case near-absolute, moral convergence. But this is not because Americans converged in their comprehensive moral worldviews. There is plenty of moral diversity swirling around the core targets of agreement." (2024, 239-240)

discussion are paramount. Philosopher Jürgen Habermas articulates a similar view about the nature of consensus-building when he states:

> To argue is to contradict. But it is only the right – and, indeed, the encouragement – to say 'no' to each other that elicits the epistemic potential of language without which we could not *learn from one another*. And this is the point of deliberative politics, namely, that by engaging in political disputes, we improve our beliefs and thereby approach the correct solution to problems. (2023, 64)

It is clear why deliberation ought to be prioritized during the process of consensus-building. If people are afraid to question or criticize ideas, then societies may be more likely to reach consensuses around ideas that are pernicious or downright wrong.[181] If a consensus is going to be reached, it ought to be reached via a process of intensive debate rather than a process wherein people are intimidated into acceptance of an idea through social or political pressure. Philosopher John Peter DiIulio, drawing on the work of Ronald Terchek, articulates Mill's approach to consensus, stating: "… Mill's drive for moral consensus on issues like 'gender inequality' is intended to be the product of the exact kind of arguments and testimony that characterizes any good, liberal politics: 'Mill believes that society could 'advance' to moral agreements about important matters, but the new consensus does not mean that it cannot be disputed.'" (2022, 271-272) With these points in mind, it should be clear why there is no contradiction or incoherence in those who wish to cultivate a Millian atmosphere of free expression simultaneously dedicating time and energy towards the achievement of societal consensus around key issues.

[181] David Zweig draws on work by philosopher of science Eric Winsberg, who notes that the punishment of heterodox thinkers can generate a "manufactured" consensus: "'Sometimes, a scientific consensus is established because vested interests have diligently and purposefully transformed a situation of profound uncertainty into one in which there appears to be overwhelming evidence for what becomes the consensus view,' wrote Eric Winsberg in an article on the necessity of scientific dissidents. A key lesson here is that consensus is often manufactured, knowingly by some, unwittingly by others. We must always look for evidence, rather than expert opinion or the appearance of consensus. 'This doesn't mean we should believe every heterodox thinker that comes along. But it means we should strongly resist the urge to punish them, to censor them, to call them racist.' Instead, Winsberg wrote, we should simply 'evaluate their claims.'" (2025, 343)

Moreover, it also seems empirically wrong to suggest that when societies reach a consensus about an issue, this necessarily entails a reduction in intellectual diversity. This is because when people resolve an issue and no longer view it as an interesting area for debate, they do not simply stop thinking and communicating with one another. Rather, they identify new areas of inquiry that are worthy of attention and energy, and begin involving themselves in debates about these matters, thereby expanding their intellectual horizons. This fact is fairly obvious and commonsensical, but it is worth spelling out in detail. While today we have achieved a broad consensus about an array of questions, it does not follow from this that productive discourse has been diminished. Instead, contemporary discourse simply revolves around issues that are more salient at this particular moment in history. Rather than eliminating intellectual diversity, the achievement of consensus enables humans to shift their attention and energy towards other issues that are more relevant and pressing for the historical epoch that they inhabit. If consensuses about such contemporary issues are eventually reached, people will simply move on to a new set of issues, and a diverse array of perspectives will be needed in these areas as well. This cyclical dynamic underscores the value in conceptualizing the process of consensus-building as part of long-term human and societal progress.

It is worth briefly noting that the project of consensus-building that has been described above is also relevant to the issue of institutional credibility, which was the focus of the previous chapter. We have seen that when epistemic institutions lose credibility, societies become increasingly fragmented, and people with different worldviews become more likely to filter out challenging information, and simply expose themselves to ideas and arguments that they find

congenial. [182] This division makes it much more difficult, or even impossible, for the project of consensus-building to take place. Intellectual diversity is a mechanism that enables humans to refine their worldview and apprehend truths, and once a truth is apprehended, it is desirable and beneficial for diverse individuals and groups to gradually coalesce around it and reach a broad agreement. Credible epistemic institutions that are heterodox in character provide an avenue towards the kind of societal consensus that Mill, among many others, views as highly valuable. Indeed, it is difficult to imagine how a strong consensus could be reached in the absence of institutions that have earned the trust of many diverse segments of society. If we accept the Millian precept that achievement of a consensus can be a major gain for society, then we have strong grounds upon which we may advocate for the establishment of institutions that facilitate free and open discourse, and are hospitable towards a vast array of different ideas and worldviews.[183]

vii.iii Robust Consensus and Illusory Consensus

The preceding discussion of the Millian project of consensus-building is relevant to the contemporary phenomenon of online intimidation culture, which is the catalyst for this entire discussion. We have seen that social media and its divisive influence can cause communication between groups with different worldviews to deteriorate, and it should come as no surprise that this undermines society's ability to reach consensus. DiResta offers a clear summary of the ways in which modern media undermines a shared sense of reality and makes it virtually impossible to achieve consensus:

> Societies require consensus to function. Yet consensus today seems increasingly impossible. Polarizing topics are black-and-white and compromise unthinkable. Our

[182] Edmans describes people's willingness to avoid amplifying information that is at odds with their own views. (2024, 231)

[183] Hopkins and Grossman explain that a refusal to provide space for dissent and intellectual diversity can cause institutions to fall into disarray. (2024, 196)

> political leadership is gridlocked. Our media feels toxic. And social media seems like a gladiatorial arena, a mess of vitriol. The culture war is everywhere. … we feel we are no longer able to speak with friends and family or to trust institutions. It increasingly seems like we don't live in the same reality. And that's because, in a very critical way, we don't. Consensus reality—our broad, shared understanding of what is real and true—has shattered … A deluge of content, sorted by incentivized algorithms and shared instantaneously between aligned believers, has enabled us to immerse ourselves in environments tailored to our own beliefs and populated with our own preferred facts. (2024, 21-22)

This author is correct in arguing that the dynamics of social media can be directly at odds with the establishment of consensus. Countless examples of online antagonism and strife make this clear. However, it is worthwhile to point out that online intimidation and the conformist pressures that it entails can, at least theoretically, be used as a tool for advancing consensus. Rather than getting large populations to arrive at an agreement through rigorous discussion and debate, societal actors can simply use social media platforms to dole out social punishments when people entertain ideas that conflict with orthodox views. Indeed, this dynamic is the norm in many societies that currently exist and many that have existed in the past.[184] Since achieving a consensus through argumentation is often very challenging, individuals and groups can attempt to expedite the process of consensus-building by intimidating people into silence and rewarding them for publicly espousing ideas that have been deemed correct by societal actors that outrank them in terms of power and influence.[185] If we cannot convince people that our ideas are sound, we can instead use proverbial carrots and sticks to keep them in a state of submission wherein they refrain

[184] Peter MacKinnon states: "What is particularly troubling today is the co-option of officialdom into the repression of speech – a cultural phenomenon that reflects the censorious and judgmental era in which we live … we should remind ourselves of history's lessons that repression – and authoritarianism – may begin slowly and spread quickly." (2024, 87)

[185] Galea offers a helpful juxtaposition between demagoguery and consensus-building: "Societal divisions have always existed, but a responsible leader, working within the liberal order, will aim to bridge these divides, or at least not to inflame them. Unfortunately, some will choose to do the opposite, calculating that the path of the demagogue is a quicker route to prominence than that of the measured consensus builder. There is a long tradition of such figures in the United States, and they have thrived in recent years, with the internet making it easier than ever to cultivate and monetize large followings based on whatever compels attention, even at the expense of the public good." (2023, 343)

from challenging our ideas. As Russell and Patterson correctly point out: "Indirect censorship can distort the types of views that are most accessible, giving people a warped perspective of consensus, and it can have a chilling effect, scaring people away from exploring or sharing their views because of fear of being accused of misinformation." (2025, 168) Rather than constructing a consensus that is supported by evidence and argumentation, people may construct a consensus that is supported by fear.

We have established that intimidation, including intimidation facilitated by social media, can be deployed towards the goal of forging a consensus around an issue. However, there are strong reasons for thinking that this mode of consensus-building is deeply flawed. A fundamental problem with this strategy is that it is unlikely to result in the achievement of a robust consensus that will prove durable and resilient over the course of generations.[186] Instead, it is likely to produce a consensus that is frail, or illusory.[187] A consensus that is held in place by fear is one that is likely to fall apart once this fear is mitigated. Let us imagine what will unfold if and when members of a society become aware that they may challenge an ostensible consensus that has been propped up by intimidation for a period of time. It seems rather obvious that a consensus of this kind is liable to be corroded as soon as people sense that they have an opportunity to distance themselves from the relevant orthodoxies without facing punishment. If fear is removed from the equation, then so

[186] Galea states: "When we fall into the gravitational pull of a consensus and do not think for ourselves, we are vulnerable to missing the reality of what we are discussing. When this reality asserts itself, it can do much to break a consensus that is not based on practical engagement with the world." (2023, 130)

[187] Thomas Chatterton Williams posits a direct connection between cancel culture and the establishment of "fake consensus": "cancellation operates with the logic and velocity of a sucker punch: the target cannot protect herself and won't even know where the attack is coming from until it has already landed. When it is effective…it results in a coercive and widespread *onlooker effect*, enforcing a fake consensus, which, ironically, functions less as a democratizing force than as an elite gatekeeping etiquette… (2025, 210)

too is the incentive to adhere to the prevailing view. Let us not forget a passage from Mill that was referenced elsewhere in this discussion:

> Our merely social intolerance kills no one, roots out no opinions, but induces men to disguise them, or to abstain from any active effort for their diffusion. With us, heretical opinions do not perceptibly gain, or even lose, ground in each decade or generation; they never blaze out far and wide, but continue to smoulder in the narrow circles of thinking and studious persons among whom they originate, without ever lighting up the general affairs of mankind with either a true or a deceptive light. And thus is kept up a state of things very satisfactory to some minds, because, without the unpleasant process of fining or imprisoning anybody, it maintains all prevailing opinions outwardly undisturbed, while it does not absolutely interdict the exercise of reason by dissentients afflicted with the malady of thought. A convenient plan for having peace in the intellectual world, and keeping all things going on therein very much as they do already. (2015, 33)

It seems that the state of affairs that Mill describes in this passage can accurately be described as one wherein an illusory consensus is present. Mill invokes a state of affairs in which dissidents throughout society continue privately doubting ideas and norms that are culturally dominant, but keep these doubts to themselves for the sake of maintaining "peace in the intellectual world". In a situation like this, actors of various kinds refrain from expressing themselves openly and engaging in substantive discussion or debate about controversial matters. It is interesting that while Mill on the one hand espouses the view that consensus-building is an extremely important societal project, he nonetheless disapproves of social dynamics wherein people engage in self-censorship and conformity for the sake of avoiding friction between themselves and other members of their society.[188] It is evident that Mill wants agreement to be reached about contentious matters, but he does not want this agreement to be the result of fear and the disingenuity that it produces.

[188] Galea offers a warning about groupthink and advocates for greater questioning of consensuses within scientific communities: "… science has a weakness for groupthink, for being swayed by the consensus simply because it is the consensus. If this is so, then we have a responsibility not just to be on guard against this tendency, but also to maintain a healthy level of iconoclasm, an instinct for pushing against the consensus as a means of testing our assumptions and ensuring that we are indeed thinking for ourselves." (2023, 126-127)

For all of these reasons, it is appropriate to make a conceptual distinction between robust consensus and illusory consensus. A robust consensus is the product of intense deliberation, and it is likely to be maintained over the course of long periods of time. In contrast, an illusory consensus is the product of intimidation, and is likely to disintegrate if and when cultural forces are realigned and people begin to feel more comfortable voicing their genuine views. Anyone who finds this view about robust consensus and illusory consensus plausible has good reason to direct their attention towards the phenomenon of online intimidation culture. We have seen repeatedly that this phenomenon can generate great chilling effects, and pressure people into conformity. This can result in the establishment of an illusory consensus as people who harbour doubts about orthodox ideas fall silent, and begin to feel increasingly isolated. In a highly polarized society wherein conflict and strife are rampant, people can feel pressured into conforming with the norms and dictates of whichever ideological camp they happen to be aligned with. Rauch invokes the work of sociologist Elisabeth Noelle-Neuman in order to describe this pattern of behaviour:

> In a manipulated or repressive social environment, people who follow the cues around them will misread the distribution of opinion. The person who believes herself to be in the minority will assume that her views are losing ground. The more isolated she feels, the less inclined she will be to express her view, and the more pressure she will feel to conform. (2021, 195)

Elsewhere in this text, Rauch states: "By swarming social media platforms and using software to impersonate masses of people, trolls can spoof our consensus detectors to create the impression that some marginal belief held by practically no one is broadly shared." (2021, 168) Here, Rauch effectively states that social media platforms can be deployed with the specific intention of constructing an illusion of consensus. It is clear that modern channels of communication can be used for the purpose of making particular views seem much more dominant than they really are, which can fuel intimidation and help erect an illusory consensus. The question

that we will explore next is whether an alternative social media ecosystem that is less toxic and more prosocial in character could actually play a role in advancing a consensus that is not illusory and prone to disintegration, but genuine and robust. While there is no clear indication that a new social media status quo will be established in the near future, it is nonetheless valuable from a philosophical perspective to consider how new media might shape the Millian project of consensus-building in beneficial ways. In the interest of balance, it is appropriate to consider social media's potential to advance genuine consensus-building in addition to its potential to generate false and brittle consensus via intimidation.

vii.iv: Social Media, Consensus-Building, and Social Goods

It should come as no surprise that social media platforms can be used in order to establish illusory consensuses that are likely to fall apart as various cultural tides continue to ebb and flow. However, it would be an error to jump to the conclusion that social media is inherently harmful to the Millian project of robust consensus-building. It is worth taking a moment to consider what might be achieved if we were to successfully cultivate a media ecosystem wherein intimidation is minimal and people in societies all across the world are afforded the opportunity to communicate with one another in a manner that is more open and dialogic. My contention is that despite its many flaws, social media can help accelerate the kind of consensus-building for which Mill advocates.[189] The fact that social media facilitates unlimited conversation across vast geographical distances means that consensuses that are reached on a small scale can rapidly expand into consensuses that take hold across countries, regions, and eventually, the globe. The desirability of a global

[189] Jacob Mchangama describes the advent of a platform that is specifically designed to facilitate consensus: "…unlike Facebook, Twitter, and YouTube, Polis is built to promote consensus and agreement rather than division and outrage. This promising precedent has been used as the basis for a dozen laws and regulations already passed in Taiwan, and has also been used by the government of Singapore and to inform local politics in the UK and the US." (2025, 379)

consensus about this or that particular issue is beside the point here. So long as we accept the premise, as Mill does, that it is desirable for humans to reach a broad consensus about certain matters, then we have solid grounds upon which we may conceptualize social media as a potential asset towards this goal.

Since social media platforms as they currently exist do not at all approximate a Millian intellectual marketplace, it is reasonable to ask what would need to change in the realm of social media in order for these services to become venues wherein people with all kinds of worldviews can come together to sort out their differences, and eventually move in the direction of a robust consensus. There are countless ways in which social media platforms can be designed and reformed, and accordingly, we are in need of some guiding principles that can assist us in cultivating a social media ecosystem that is conducive to consensus-building rather than conflict. The three social goods that have been highlighted throughout this discussion can function as guiding principles of this kind. Since Chapter 3 offered a detailed discussion of the ways in which social media can damage free expression and its associated social goods, we will now attempt to do the opposite: we will consider how this form of technology may bolster expressive freedom and the important social goods that accompany it.

Of course, the goal here is not to provide predictions about how the social media sector will evolve over time. Nobody can know with certainty what the future holds for this relatively new form of communication. Rather, the goal here is to provide a more detailed account of the prosocial promise of social media by examining how this form of technology can be used to advance social goods rather than thwart them. While social media has played a role in undermining

expressive freedom, in addition to critical intellectual faculties, authenticity in discourse, and equity in accountability, this troubling pattern can be reversed if appropriate design choices are made and the incentives of social media discourse are arrayed in a better direction.[190] Rather than eroding these social goods, social media can be used in order to bolster them, and use them to facilitate greater understanding between diverse individuals and groups about a range of contentious issues.

In order to arrive at a robust consensus, people must be willing to be exposed to the many different views that can be offered about an issue, and give them a fair and charitable hearing. The process of examining and assessing ideas in a composed and reasoned manner necessarily involves the deployment of critical intellectual faculties. If people lack the vocabulary and conceptual tools needed to parse arguments and determine whether their premises offer adequate support for their conclusions, then the practice of assessing ideas will be much more likely to degenerate into personal skirmishes that fail to advance mutual understanding and the project of consensus-building. Accordingly, it has been argued that online intimidation culture undermines the development of critical intellectual faculties by shutting down discourse and preventing people from experimenting with different ideas in a freewheeling fashion. If the problem of online intimidation culture were to be successfully overcome, then social media platforms could have the opposite effect on discourse. Rather than aligning themselves with an ideological camp and reaping rewards for demonstrating their loyalty to this camp, social media users could be provided

[190] Rose-Stockwell considers how a social media algorithm might be designed with the aim of facilitating productive discourse: "It could show the best version of opposing positions on controversial topics. It might work to facilitate consensus on hard but necessary moral actions by offering the best version of the opposing side's arguments on every contentious issue." (2023, 361)

with an information environment wherein they would have opportunities to interact with people who are very different from themselves in myriad ways, and also have opportunities to engage them in rigorous debate.

This is important with respect to critical intellectual faculties because in a renewed information environment that is more dialogic than the one that exists at present, people could very well find themselves in a situation wherein they must think carefully about ideas that they had long simply taken for granted in order to respond to challenges and criticisms coming from others. In some cases, individuals might successfully formulate responses that are capable of answering such objections. In other cases, they might determine that their long-held views are flawed, and must accordingly be revised in some way. In many cases, people may simply feel less certain about their own positions, and reach the conclusion that a given issue is more complex and multifaceted than they had previously realized. Interlocutors engaged in public discourse might feel increasingly agnostic about a particular question as they come to realize that individuals and groups with many different perspectives actually have something of value to say about it. All of these outcomes are desirable with respect to the social good of critical intellectual faculties. This is because they involve people being required to examine the strengths and weaknesses of an array of views, and provide others with reasons if they wish to advance acceptance of their own views. This is much more likely to spur intellectual development than the tribal politics that have pervaded social media throughout recent history. In the context of such politics, people are often punished by their own ideological camp for asking difficult questions and giving a fair hearing to alternative perspectives, which stifles intellectual growth and maturation.

If we are able to successfully cultivate a social media ecosystem that functions as an engine of reasoned deliberation rather than intimidation, then this could also have desirable implications with respect to the social good of authenticity in discourse. We have seen that online intimidation can undermine authenticity in discourse by pressuring large populations into self-censorship, or even outright preference falsification. It is plain to see how this can result in the generation of an illusory consensus as people respond to social incentives by misrepresenting their own views, thereby contributing to an impression that their community has coalesced around particular ideas, when this is in fact not the case. The more that people must live in fear that their expressive acts will be met with social punishment, the more likely they will be to find themselves acting as participants in an illusory consensus.[191] Messina offers a relevant observation about the impact of aggressive and accusatory speech on people's beliefs:

> … there is reason to worry that uncivil rhetoric is not especially likely to change the beliefs of those who are targeted by it. Although uncivil speech is highly pleasing to those who engage in it and to those antecedently inclined to agree with the speaker's message (allowing them to delight in the feeling of righteous indignation), it remains disagreeable to those who do not. I may employ all sorts of rhetorical tools and logical fallacies to convince you of something you have no reason to believe. But I am not likely to be able to persuade [you] to join my side … by yelling at you, calling you names, and refusing to take you at your word. If I'm lucky, I may succeed in cowing you into self-censorship and presenting a false front. But that's compatible with me going on believing just as before. (2023, 48)

Messina is surely correct that there is a large gap between intimidating people into silence, and successfully getting people to accept one's views. What is interesting for our purposes is that to external observers, it may be difficult or impossible to tell the difference between the two, and

[191] Daniel F. Stone also highlights the risk of incorrectly perceiving that a consensus is present: "As we encounter bad actors on the other side online and in conversation more often, we'll thus be more likely to overestimate their general prevalence … Social pressure can make people hesitant to speak up when they think out-partisans are being characterized unfairly. Limited strategic thinking and selection neglect can again make us fail to account for the absence of these dissenting voices and be overly influenced by superficial consensus. (2023, 141)

that is why the distinction between illusory consensus and robust consensus is important. In a state of affairs wherein such fears are mitigated or eliminated, social media could have a very different effect on authenticity in discourse than the one described above. Social media, practically by definition, has appeal because it is a venue wherein nearly anyone can create an account and choose to participate. This inclusive design can have benefits as well as drawbacks, but it is obvious enough that low barriers to entry are what make social media fundamentally different from other kinds of media such as newspapers, magazines, television, films, etc. The absence (or near-absence) of gatekeepers in the realm of social media provides space for countless individuals and groups to express themselves and attempt to find an audience who will be receptive to what they have to say.

This openness is particularly significant in situations wherein marginalized individuals and groups find that they are either ignored or misrepresented by more traditional media institutions, which may even have financial reasons for ensuring that their voices are not amplified or elevated. For example, let us imagine a scenario wherein employees of a powerful corporation feel that they are being exploited or mistreated, but struggle to have their concerns documented by news outlets that rely on said corporation for advertising revenue.[192] Social media platforms provide an alternative venue for the dissemination of such information and ideas that can enable people to circumvent individuals and institutions that would like to function as gatekeepers. This can lead to greater authenticity in discourse by enabling people with very little social, economic, or

[192] Taibbi describes how commercial pressures can drive journalists away from challenging the practices of powerful businesses: "The biggest outlets learned there's no percentage in doing big exposés against large, litigious companies. Not only will they sue, but they're also certain to pull ads as punishment … The message to reporters working in big corporate news organizations was that long-form investigative reports targeting big commercial interests weren't forbidden exactly, just not something your boss was likely to gush over." (2021, 76)

institutional power to speak candidly about their views and concerns without intermediaries shaping their messages and influencing the manner in which they will be received.

If it is indeed true that social media can promote authenticity in discourse, then this will have implications for the Millian project of consensus-building. If we successfully establish a system of online communications that can unite people from many different countries, cultures, socioeconomic backgrounds, and the like so they may engage in dialogue that is sincere and uncorrupted by intimidation, then this can bolster efforts to achieve agreement about contentious matters. While there is of course no guarantee that an atmosphere of free expression in the realm of social media will lead to the achievement of a robust consensus, it does mean that if and when consensus is achieved, it is likely to be robust rather than illusory. This is because such a consensus will be the result of discussion and argumentation rather than self-censorship and preference falsification. It would be erroneous to suggest that greater authenticity in discourse leads everywhere and always to the achievement of consensus, but it is reasonable to point out that if and when we sense that a new consensus is emerging, authenticity in discourse can give us confidence that this consensus is a real and credible one that will have the ability to remain resilient over the long term.

Let us now consider the social good of equity in accountability. It was noted early on in this work that the dynamics of social media make it remarkably easy for users to engage with high-profile people who have amassed large followings, regardless of whether or not the users in question have amassed a sizeable following themselves. In many cases, this relationship can turn toxic as people deploy attacks on high-profile individuals as a means of growing their own social

media following and boosting their relevance on a given platform. While these dynamics are pernicious for a variety of reasons, it is worth noting that the inclusion of ordinary people on social media platforms alongside others who are exceptionally influential can have positive ramifications with respect to the social good of equity in accountability. This is because by providing a venue wherein people with elite status are encouraged to interact with others who do not possess such a status, social media platforms make it possible for the words and ideas of various elites to receive scrutiny and feedback that they might never receive otherwise.

A theme that has informed previous chapters is the harmful character of siloing, or division between different social groups.[193] Since siloing undermines communication and understanding between individuals and groups with different worldviews,[194] it is necessarily antithetical to the atmosphere of free expression that Mill, and the many philosophers and theorists that he has influenced, wish to promote.[195] Ross provides a clear summary of the issues associated with siloing:

> The call out culture means you get to discriminate in favor of those who agree with you. But tribalism is still tribalism, whether Left or Right—and a call out culture makes our tribes smaller and more impotent. Loyalty to the tribe becomes more important than coexisting peacefully with others in a pluralistic system. And as millennial journalist Malikia Johnson pointed out, 'being encapsulated within silos of their own thoughts' causes people 'to mistakenly think that a larger part of the world agrees with their points of understanding.' This part of cancel culture should be canceled. What is missing in our

[193] Philosopher Peter Ives states that "…Tim Berners-Lee, one of the developers of the standard protocols that made the World Wide Web possible in the 1990s, has similar worries about the dysfunctional nature of current social media 'silos.'" (2024, 120)

[194] McIntyre offers some advice about how people ought to resist siloing and its associated distrust: "Even if you are on the virtuous side of facts and truth, fragmentation is dangerous. Remember that the goal of a disinformation campaign is not merely to get you to doubt, but also to distrust anyone on the other side … As hard as it is, do not merely retreat to your silo and 'be right.' Reach out to those who disagree with you, who have been misinformed and disinformed. If at all possible, try to do so with kindness. They do not need another person to hate or distrust." (2023, 121-122)

[195] Jonathan Turley uses the language of siloing in his criticism of post-secondary educational institutions: "Academia will likely remain a battleground over the meaning of free speech, since faculties show little evidence that they will yield to calls for greater diversity of thought and expression. Despite stinging losses in the courts, colleges and universities remain a hardened silo of speech intolerance." (2023, 307-308)

> distorted debate about cancel culture is that calling out is a powerful tool, but it isn't always the right tool for the job. Even when a call out is justified, it's not always productive. (2025, 48-49)

In light of concerns about siloing, it is reasonable to posit that one of the potential benefits of a renewed social media ecosystem that is not pervaded by intimidation is that it will provide a space wherein ideas espoused by powerful and influential segments of society can be checked by others who possess less in terms of power and influence. In other words, social media platforms can function as spaces wherein powerful societal actors are called to account. Importantly, calling one to account in this context does not involve ad hominem attacks, ostracism, or anything of the like. Rather, it simply involves influential figures being required to explain and defend their views when these views are challenged by the general public.[196] Social media can bolster equity in accountability by elevating the likelihood that powerful people and institutions will not be able to insulate themselves from criticism when they present their ideas to the world.

It is common for people to criticize politicians and other influential figures who shun media outlets that are critical towards them, and choose to exclusively give interviews to people and institutions that will be friendly or flattering towards them. This is a form of siloing wherein powerful figures can enjoy comfort and praise as they interact with populations that already support them, and simply avoid populations that might pose difficult questions or raise objections towards their ideas, behaviour, and policies. Fortunately, this kind of insularity can be challenged

[196] Patterson and Russel offer relevant commentary about the need for elite consensuses to be open to criticism and revision: "A particular problem with the kind of 'expertise' that … elites celebrate is that it is blind to the quality of expertise …The solution to this problem is to do what academia has always done: engage in debate, dissent, discussion, repeated testing, and eventually consensus—but always a tentative consensus that in turn gives way to more debate, dissent, and revision of theories. Alas, this is not the kind of expertise that elites seem to have in mind when they tell us to 'defer to experts.' Expertise is a process, but they want to define it as an outcome." (2025, 83-84)

via social media platforms that bring together vast collections of voices. A social media ecosystem that successfully achieves a culture of intellectual diversity and openness can help put a stop to siloing by placing elites in an information environment wherein hiding from scrutiny is difficult, or perhaps even impossible. If we can reverse the trend of social media users sorting themselves into insular communities wherein they are rewarded for their loyalty, and create an online environment wherein many different people feel comfortable expressing heterodox views, then we can bolster equity in accountability by ensuring that influential persons, and the large audiences that follow them, are confronted with alternative perspectives. Most ordinary people must face criticism on a routine basis and answer for their mistakes and shortcomings, and there is no good reason why people in positions of power should be shielded from this process. The social good of equity in accountability will likely be bolstered in media environments wherein elites receive feedback from supporters and critics alike, rather than simply being surrounded by people and institutions that are committed to protecting them.

There is no doubt that the language of equity in accountability may strike many as lofty. However, if we carefully examine the contemporary political landscape and note the willingness of elite persons to distance themselves from individuals and institutions that may call them to account in a serious way, then we can develop a clearer understanding of the importance of equity in accountability for liberalism as well as democracy.[197] This is a social good that ought to be

[197] This objective is particularly salient given the apparent willingness of the 47th president of the United Status and his administration to punish media organizations that they perceive as hostile, and to reward those that they perceive as friendly or loyal. In 2025, FIRE offered the following statement about the president's decision to block Associated Press reporters from events at the White House due its refusal to use his preferred language in its reporting: "Punishing journalists for not adopting state-mandated terminology is an alarming attack on press freedom … President Trump has the authority to change how the U.S. government refers to the Gulf. But he cannot punish a news organization for using another term. The role of our free press is to hold those in power accountable, not to act as their mouthpiece." (Siemaszko 2025)

actively cultivated, and my view is that despite the many failings of social media companies over the course of recent history, this channel of communication can nonetheless play a role in ensuring that the powerful are scrutinized and checked in appropriate ways. A media ecosystem wherein elites can pick and choose which outlets to deal with in the interest of rewarding partisan loyalty, protecting their image, and evading meaningful debate, is one wherein discourse will become impoverished, and politics will likely become increasingly dysfunctional. We thus have a slew of good reasons for combating siloing of the media ecosystem and ensuring that elites are not permitted to evade accountability by inhabiting a media silo that exists in order to promote them and augment their power.

vii.v: Social Media and the Expansion of Compassion

We have seen that a renewed social media ecosystem could have positive implications with respect to the project of building consensuses between diverse individuals and groups around the world. At this juncture, I would like to deepen this account of the prosocial potential of social media by examining the subject of compassion. The idea that we will consider is whether in addition to facilitating communication between diverse populations, social media can also facilitate compassion between them. Let us take a moment to consider Mill's views about the topic of compassion, or what he refers to as "sympathy". Simply put, Mill thinks that individuals and communities have an enormous capacity for compassion and caring, and even indicates that this capacity can extend to non-human organisms. Consider the following passage from *Utilitarianism*:

> It is natural to resent, and to repel or retaliate, any harm done or attempted against ourselves, or against those with whom we sympathize … Whether it be an instinct or a result of intelligence, it is, we know, common to all animal nature; for every animal tries to hurt those who have hurt, or who it thinks are about to hurt, itself or its young. Human beings, on this point, only differ from other animals in two particulars. First, in being capable of sympathizing, not solely with their offspring, or, like some of the more noble

> animals, with some superior animal who is kind to them, but with all human, and even with all sentient, beings. Secondly, in having a more developed intelligence, which gives a wider range to the whole of their sentiments, whether self-regarding or sympathetic. By virtue of his superior intelligence, even apart from his superior range of sympathy, a human being is capable of apprehending a community of interest between himself and the human society of which he forms a part, such that any conduct which threatens the security of the society generally, is threatening to his own, and calls forth his instinct (if instinct it be) of self-defence. The same superiority of intelligence, joined to the power of sympathizing with human beings generally, enables him to attach himself to the collective idea of his tribe, his country, or mankind, in such a manner that any act hurtful to them rouses his instinct of sympathy, and urges him to resistance. (2015, 164-165)

Mill's reference to people engaging in solidarity with their "tribe…country, or mankind" suggests that humans' capacity for compassion and caring can grow in terms of its scope and strength as human societies evolve and become more sophisticated. While people may have sympathies that are relatively narrow at certain points in time, these sympathies can be expanded given appropriate environmental conditions. This sensibility is highly congruent with Mill's optimistic view of the growth and maturation that can be achieved through the deployment of reason. Since humans are capable of spurring progress through the deployment of their many cognitive and physical gifts, there is no discernable limit on just how far the sympathies that Mill describes might extend. While Mill does not offer a comprehensive analysis of communications technology in his writings, his bold arguments about the human capacity for compassion can be helpful for understanding why innovation in this area can plausibly be viewed as an asset to human cooperation and flourishing in many cases. As the ability of people to communicate with others becomes strengthened, so too does the ability of said people to become compassionate, or sympathetic, towards individuals and groups that may have seemed entirely foreign or alien to them at one point in time.

It is not too difficult to see how technological advancement might shape people's attitudes towards geographically distant individuals and groups that are different from themselves in terms of language, culture, ethnicity, and so on. It is one thing to have some vague awareness of people living in faraway lands, and it is another to communicate directly with these people through forms of media such as text, audio, and video. The more that people are able to establish rapport with others living in distant societies, the more difficult it becomes for them to think of these persons as anonymous, undifferentiated strangers. While there is no doubt that innovations in communications technology can empower bad actors and generate social strife, they can also play a role in cultivating feelings of solidarity and community between people that might otherwise have no significant awareness of, or interest in, one another. The human capacity for compassion can progress alongside technological innovation, thanks to the new forms of communication and dialogue that it enables.

This is a point that is alluded to by philosopher Peter Singer in the 2011 version of his book *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Singer points out that the rise of digital media can have dramatic implications with respect to ethics. Indeed, he seems somewhat awestruck at the immense power of new media to facilitate communication and discussion between people from all walks of life, all around the planet:

> Recording our thoughts digitally, rather than on paper, means that they can be sent electronically, and the availability of instant, virtually free communication all over the world is affecting every aspect of our lives, including our ethics … I quote Gunnar Myrdal's *An American Dilemma*, a major study of attitudes about race and racism published in 1944. In Myrdal's view, greater social mobility, more intellectual communication, and more public discussion were already then contributing to a change in the racist attitudes that had existed for so long in some parts of the United States. If more mobility and more communication were already making a difference in 1944, what should we expect from the vastly greater changes that are happening now, linking people all over the world, and opening up communities that hitherto had little access to ideas from outside?

> The experiment is under way, and there will be no stopping it. What it will do for the rate at which we make moral progress and expand the circle of those about whom we are concerned, remains to be seen. (2011, xiii - xiv)

Clearly, Singer understands that technological progress can be highly relevant with respect to ethics. As a staunch advocate for animal rights as well as other social causes, Singer consistently promotes the notion that the human capacity for compassion ought to be expanded, and that a core component of human progress consists in the inclusion of more living beings in our circle of moral concern. In a passage that appears to channel Millian sensibilities, he states:

> The circle of altruism has broadened from the family and tribe to the nation and race, and we are beginning to recognize that our obligations extend to all human beings. The process should not stop there … The only justifiable stopping place for the expansion of altruism is the point at which all whose welfare can be affected by our actions are included within the circle of altruism. This means that all beings with the capacity to feel pleasure or pain should be included; we can improve their welfare by increasing their pleasures and diminishing their pains. (2011, 120)

It is easy to see why the evolution of online communication that has unfolded over the last few decades could play a role in broadening the "circle of altruism" that is at the heart of Singer's philosophy. It is not only the case that people can now exchange text messages with people in faraway places; they can also participate in high-quality audio and video communications in real time. When people experience hardships or other notable events, they can document them with smartphones and share these experiences via social media for audiences consisting of thousands, or even millions. While major events such as wars, terrorist attacks, and police violence towards civilians may have once seemed abstract to people who were learning about them, these events are likely to seem much more palpable when they are thoroughly documented through new media. This documentation also makes them more difficult to ignore. There is no doubt that exposure to such content can have some undesirable consequences, such as stoking fear and anxiety among

populations who have virtually endless access to news media and the disturbing information that it includes. However, it can also have the highly desirable effect of causing people to view others, who may be different from themselves in many respects, as individuals who are worthy of moral concern and the compassion that comes with it.

Many will notice a similarity between Mill and Singer's writings. Singer's view that humans' circle of moral concern should extend to all organisms that have the ability to experience pleasure and pain is strikingly similar to Mill's view, noted above, about the capacity for humans to extend their sympathies towards all sentient beings. Despite the fact that Singer does not make reference to Mill in this particular text, it is clearly the case that there is overlap between Singer's arguments and Mill's arguments about the impressive ability of humans and their communities to drive moral progress by including more and more subjects in their domain of ethical consideration. Moreover, Singer's invocation of the importance of "public discussion" is highly congruent with the Millian philosophy of free expression that has animated this dissertation. Singer endorses the Millian idea that free expression can be an engine of social progress by allowing populations to consider and analyze many competing ideas about contentious topics, which prevents entrenched orthodoxies from standing in the way of positive change.

It should be clear enough why the philosophical positions staked out by Mill and Singer lend credence to the idea that social media can be a force for good. As the latter author points out, technology can play a key role in moving humanity closer to the ambitious goal of expanding its circle of moral concern and viewing many diverse beings as worthy of altruistic treatment. Accordingly, it is reasonable to argue on Millian and Singerian grounds that despite its many flaws,

social media has the potential to generate real and significant gains for many individual and groups by facilitating meaningful communication between them. The more challenging question is not whether social media is capable of producing these desirable outcomes, but whether human populations will tap into the awesome power of this technology for prosocial purposes instead of permitting it to function as an engine of conflict and extremism. The many criticisms of social media that have been articulated throughout the preceding chapters have been advanced in hopes that rather than simply giving up on the project of cultivating a healthy social media ecosystem, people will eventually harness the power of this technology in order to realize worthwhile objectives.

vii.vi: Concluding Remarks

My hope is that this chapter has provided compelling reasons to have measured optimism about the future of social media. Rather than viewing this technological innovation as a burden upon public discourse and society more broadly, we are better served by a nuanced accounting of the many benefits and drawbacks that can be associated with it. There is no contradiction in maintaining that while social media has played a critical role in corroding public discourse over the course of the 2010s and 2020s, it may well prove to be very beneficial to public discourse in the future. The introductory chapter of this dissertation used the incentives that are present in social media discourse as a starting point for understanding the many social ills that can be generated or exacerbated by it. The turbulent and dysfunctional character of online communication can be much easier to understand and critique once we develop an understanding of how incentives shape people's behaviour and their interactions with others. Rather than examining our modern media ecosystem and reaching the conclusion that social media ought to be avoided, it is more productive to contemplate how it could evolve if it were animated and propelled by a different set of

incentives. We can and should be highly critical of social media companies and the leaders who govern them, while also striving to maximize the benefits that can be realized through these platforms.

Before concluding this discussion, I would like to address a concern that might arise with respect to the arguments presented in this chapter, as well as all of the chapters that have preceded it. Even if one agrees with the arguments about intimidation culture, social media, and free expression that have been offered, they may nonetheless by irked by a sense that the chilling effects that have propelled this entire project are really just the product of a particular moment in history, and that accordingly, it is inappropriate to dedicate so much time and energy to discussing this issue and potential remedies. Some might raise the criticism that while intimidation culture is a pernicious phenomenon, it is also a transient phenomenon, and that there is little need to think in such detail about an issue that is likely to recede with time.

It is undoubtedly true that culture is constantly evolving and being challenged, and that it is unlikely that the stifling of discourse facilitated by social media will continue uninterrupted over the long term. Indeed, it seems that this would be nearly impossible given the role that younger generations play in shaping culture and challenging entrenched patterns of thinking and behaviour throughout society and its institutions. It is entirely possible that the stifling impact of social media on intellectual and expressive freedom will abate in the future, or that this is already taking place as societies move on from the particular issues and controversies that pervaded public discourse throughout the 2010s and early 2020s. Since cultural norms are so prone to fluctuation, I will offer no predictions about the prevalence of self-censorship and conformist pressures over the coming

years and decades. Too many variables can play a role in these phenomena for anyone to have confidence about their future trajectory. Instead, I will simply use this opportunity to advance the modest claim that even if the forces of intimidation do peak and then wane for a period of time, this does not mean that they have been defeated and will continue to fade into the past. Societies can vacillate between a culture of free expression and a culture of intimidation as years progress and various concerns enter and exit the foreground of public concern. If one comes across evidence that rates of self-censorship and its attendant conformity are in decline, which would certainly be welcome for anyone who embraces a Millian view of free expression, they should practice caution rather than rushing to the conclusion that intimidation culture is no longer a relevant topic of discussion. This is because even if intimidation culture does taper off for a period of time, it is entirely possible that it will be reinvigorated thanks to unforeseeable shifts in domains such as politics, economics, and technology.

Moreover, the particular dynamics of intimidation culture can change as different social groups compete for power and attempt to dislodge one another from their long-held positions of authority and influence. An individual who is highly concerned about the influence of activists on university campuses might be relieved if and when they observe that these actors' attempts at policing discourse appear to be losing momentum. However, instead of marking the end of intimidation culture, this change could really be a sign of a different form of intimidation gaining cultural ground as rival activists deploy their own power and influence in order to shape discourse. Anytime punishment is used to prevent voices from receiving a fair hearing, we have strong grounds to worry about intimidation undermining a healthy and productive public discourse. Since intimidation and punishment can emanate from individuals and groups of any ideological

persuasion, it is important to remain sensitive to the fact that an ostensible decline in intimidation culture could really just amount to a proverbial changing of the guard wherein the censorious tendencies of one ideological camp are overshadowed and supplanted by the censorious tendencies of a different ideological camp. My hope is that the analysis and arguments that have been presented throughout this dissertation can play a role in building a culture that is resilient in the face of intimidation attempts, regardless of which individuals and groups happen to be deploying intimidation tactics at any particular moment in history.

Bibliography

al-Gharbi, Musa. *We Have Never Been Woke: The Cultural Contradictions of a New Elite.* Princeton: Princeton University Press, 2024.

Anderson, Elizabeth. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton: Princeton University Press, 2017.

Aral, Sinan. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health - and How We Must Adapt*. New York: Currency, 2021.

Bail, Chris. *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing.* Princeton: Princeton University Press, 2022.

Banout, Tony, and Tom Ginsburg. *The Chicago Canon on Free Inquiry and Expression*. Chicago: The University of Chicago Press, 2024.

Ben-Porath, Sigal R. *Cancel Wars: How Universities Can Foster Free Speech, Promote Inclusion, and Renew Democracy.* Chicago: The University of Chicago Press, 2023.

Benkler, Yochai, Robert Faris, and Hal Roberts. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press, 2018.

Berlin, Isaiah. *Liberty*. Oxford: Oxford University Press, 2002.

Bode, Leticia, and Emily K. Vraga. *Observed Correction: How We Can All Respond to Misinformation on Social Media*. New York: Oxford University Press, 2025.

Bok, Derek. *Attacking the Elites: What Critics Get Wrong—and Right—About America's Leading Universities.* New Haven: Yale University Press, 2024.

Bollinger, Lee C., and Geoffrey R. Stone, eds. *Social Media, Freedom of Speech, and the Future of Our Democracy.* New York: Oxford University Press, 2022.

Boy, John D., and Justus Uitermark. *On Display: Instagram, the Self, and the City*. New York: Oxford University Press, 2024.

Burnett, Alycia, Devin Knighton, and Christopher Wilson. "The Self-Censoring Majority: How Political Identity and Ideology Impacts Willingness to Self-Censor and Fear of Isolation in the United States". *Social Media + Society*. July – September 2022.

Carlson, Taylor N., and Jaime E. Settle. *What Goes Without Saying: Navigating Political Discussion in America*. Cambridge: Cambridge University Press, 2023.

Caulfield, Mike, and Sam Wineburg. *Verified: How to Think Straight, Get Duped Less, and Make Better Decisions about What to Believe Online.* Chicago: The University of Chicago Press, 2023.

Chong, Joshua. "Beer sales hit all-time low as Canada's alcohol sales see largest drop in a decade, new report finds". *Toronto Star*. February 27, 2023.

Christensen, Jen. "US cigarette smoking rate falls to historic low, but e-cigarette use keeps climbing". *CNN*. April 27, 2023.

Corn-Revere, Robert. *The Mind of the Censor and the Eye of the Beholder*. Cambridge: Cambridge University Press, 2021.

Daub, Adrian. *The Cancel Culture Panic: How an American Obsession Went Global*. Stanford: Stanford University Press, 2024.

DeNardis, Laura. *The Internet in Everything: Freedom and Security in a World with No Off Switch*. New Haven: Yale University Press, 2020.

DiIulio, John Peter. *Completely Free: The Moral and Political Vision of John Stuart Mill,* Princeton: Princeton University Press, 2022.

DiResta, Renee. *Invisible Rulers: The People Who Turn Lies into Reality.* New York: PublicAffairs, 2024.

Duffin, Jacalyn. *COVID-19: A History*. Montreal: McGill-Queen's University Press, 2022.

Edmans, Alex. *May Contain Lies: How Stories, Statistics, and Studies Exploit Our Biases—And What We Can Do about It*. Oakland: University of California Press, 2024.

Fuerstein, Michael. *Experiments in Living Together: How Democracy Drives Social Progress*. New York: Oxford University Press, 2024.

Fukuyama, Francis. *Liberalism and Its Discontents*. New York: Farrar, Straus and Giroux, 2022.

Galea, Sandro. *Within Reason: A Liberal Public Health for an Illiberal Time*. Chicago: The University of Chicago Press, 2023.

Galloway, Scott. "TikTok Debate". *Munk Debates*. April 3, 2024.

Gershberg, Zac, and Sean Illing. *The Paradox of Democracy: Free Speech, Open Media, and Perilous Persuasion.* Chicago: University of Chicago Press, 2023.

Gibson, James L. and Joseph L. Sutherland. "Keeping Your Mouth Shut: Spiraling Self-Censorship in the United States." *Political Science Quarterly*. June 9, 2023.

Glennon, Michael J.. *Free Speech and Turbulent Freedom: The Dangerous Allure of Censorship in the Digital Era*. New York: Oxford University Press, 2024.

Grafstein, Isaac. "Debate: Should the U.S. Ban TikTok?" *The Free Press*. March 22, 2024.

Grossmann, Matt, and David A. Hopkins. *Polarized by Degrees: How the Diploma Divide and the Culture War Transformed American Politics*. Cambridge: Cambridge University Press, 2024.

Haidt, Jonathan. *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness.* New York: Penguin Press, 2024.

Haidt, Jonathan. "Why the Past 10 Years of American Life Have Been Uniquely Stupid". *The Atlantic*, April 11, 2022.

Hannan, Jason. *Trolling Ourselves to Death: Democracy in the Age of Social Media.* New York: Oxford University Press, 2024.

Hare, Ivan, and James Weinstein, ed. *Extreme Speech and Democracy*. Oxford: Oxford University Press, 2009.

Harris, Tristan. "Down the Rabbit Hole by Design. Guest: Guillaume Chaslot". *Your Undivided Attention*. July 10, 2019.

Harris, Tristan. "With Great Power Comes… No Responsibility? with Yaël Eisenstat". *Your Undivided Attention*. June 25, 2019.

Ho, Benjamin. *Why Trust Matters: An Economist's Guide to the Ties That Bind Us*. New York: Columbia University Press, 2021.

Hund, Emily. *The Influencer Industry: The Quest for Authenticity on Social Media.* Princeton: Princeton University Press, 2023.

Ives, Peter. *Rethinking Free Speech.* Halifax: Fernwood Publishing, 2024.

Jarvis, Jeff. *The Web We Weave: Why We Must Reclaim the Internet from Moguls, Misanthropes, and Moral Panic*. New York: Basic Books, 2024.

Joshi, Hrishikesh. *Why It's OK to Speak Your Mind.* New York: Routledge, 2021.

Kalmoe, Nathan P., and Lilliana Mason. *Radical American Partisanship: Mapping Violent Hostility, Its Causes, and the Consequences for Democracy*. Chicago: The University of Chicago Press, 2022.

Kiros, Hana. "Hated that video? YouTube's algorithm might push you another just like it." *MIT Technology Review*. September 20, 2022.

Kligler-Vilenchik, Neta, and Ioana Literat. *Not Your Parents' Politics: Understanding Young People's Political Expression on Social Media*. New York: Oxford University Press, 2024.

Kosseff, Jeff. *The Twenty-Six Words that Created the Internet.* Ithaca: Cornell University Press, 2019.

Kuran, Timur. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge: Harvard University Press, 1997.

Langton, Rae. *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. Oxford: Oxford University Press, 2009.

Lanier, Jaron. *Ten Arguments for Deleting Your Social Media Accounts Right Now.* New York: Picador, 2018.

Lewis, Hyrum, and Verlan Lewis. *The Myth of Left and Right: How the Political Spectrum Misleads and Harms America.* New York: Oxford University Press, 2023.

Lindsay, Roddy. "To Fix Section 230, Target Algorithmic Amplification". *The Information*. October 27, 2020.

Loury, Glenn C.. *Self-Censorship*. Cambridge: Polity Press, 2025.

Lukianoff, Greg, and Jonathan Haidt. *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure.* New York: Penguin Books, 2018.

Lukianoff, Greg, and Rikki Schlott. *The Canceling of the American Mind: Cancel Culture Undermines Trust and Threatens Us All—But There Is a Solution.* New York: Simon & Schuster, 2023.

Lynch, Michael Patrick. *On Truth in Politics: Why Democracy Demands It*. Princeton: Princeton University Press, 2025.

Macedo, Stephen, and Frances Lee. *In Covid's Wake: How Our Politics Failed Us*. Princeton: Princeton University Press, 2025.

MacKinnon, Peter. *Confronting Illiberalism: A Canadian Perspective*. Toronto: University of Toronto Press, 2025.

Marietta, Morgan, and David C. Barker. *One Nation, Two Realities: Dueling Facts in American Democracy*. New York, Oxford University Press, 2019.

Marwick, Alice E. *The Private Is Political: Networked Privacy and Social Media*. New Haven: Yale University Press, 2023.

Matsuda, Mari J., Charles R. Lawrence III, Richard Delgado, and Kimberlè Williams Crenshaw. *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*. New York: Routledge, 2018.

Matthes, Erich Hatala. *Drawing the Line: What to Do with the Work of Immoral Artists from Museums to the Movies.* New York: Oxford University Press, 2021.

Mchangama, Jacob. *Free Speech: A History from Socrates to Social Media*. New York: Basic Books, 2025.

McIntyre, Lee. *On Disinformation: How to Fight for Truth and Defend Democracy*. Cambridge: The MIT Press, 2023.

McLaughlin, Sarah. *Authoritarians in the Academy: How the Internationalization of Higher Education and Borderless Censorship Threaten Free Speech.* Baltimore: Johns Hopkins University Press, 2025.

Messina, J.P. *Private Censorship*. New York: Oxford University Press, 2023.

Mill, John Stuart. *On Liberty, Utilitarianism, and Other Essays*. Edited with an Introduction and Notes by Mark Philp and Frederick Rosen. New York: Oxford University Press, 2015.

Mitts, Tamar. *Safe Havens for Hate: The Challenge of Moderating Online Extremism*. Princeton: Princeton University Press, 2025.

Mounk, Yascha. *The Identity Trap: A Story of Ideas and Power in Our Time*. New York: Penguin Press, 2023.

Mutz, Diana C. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton: Princeton University Press, 2015.

Nichols, Tom. *The Death of Expertise: The Campaign against Established Knowledge and Why it Matters*. New York: Oxford University Press, 2024.

O'Connor, Calin, and James Owen Weatherall. *The Misinformation Age: How False Beliefs Spread*. New Haven: Yale University Press, 2019.

Owen, Taylor, and Supriya Dwivedi. "Whose speech will Elon Musk's Twitter be protecting, exactly?" *The Globe and Mail*. April 27, 2022.

Petre, Caitlin. *All the News That's Fit to Click: How Metrics Are Transforming the Work of Journalists.* Princeton: Princeton University Press, 2021.

Picciolini, Christian. *Breaking Hate: Confronting the New Culture of Extremism*. New York: Hachette Books, 2020.

Radzik, Linda, Christopher Bennett, Glen Pettigrove, and George Sher. *The Ethics of Social Punishment: The Enforcement of Morality in Everyday Life*. Cambridge: Cambridge University Press, 2020.

Rauch, Jonathan. *The Constitution of Knowledge: A Defense of Truth*. Washington: Brookings Institution Press, 2021.

Redstone, Ilana, and John Villasenor. *Unassailable Ideas: How Unwritten Rules and Social Media Shape Discourse in American Higher Education.* Oxford University Press, 2020.

Redstone, Ilana. *The Certainty Trap: Why We Need to Question Ourselves More—and How We Can Judge Others Less*. Durham: Pitchstone Publishing, 2024.

Ronson, Jon. *So You've Been Publicly Shamed*. New York: Riverhead Books, 2016.

Rose-Stockwell, Tobias. *Outrage Machine: How Tech Amplifies Discontent, Disrupts Democracy—And What We Can Do About It.* New York: Legacy Lit, 2023.

Rosenblum, Nancy L. and Russell Muirhead. *A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy*. Princeton: Princeton University Press, 2019.

Ross, Loretta J. *Calling In: How to Start Making Change with Those You'd Rather Cancel.* New York: Simon & Schuster, 2025.

Russell, Jacob Hale, and Dennis Patterson. *The Weaponization of Expertise: How Elites Fuel Populism*. Cambridge: The MIT Press, 2025.

Settle, Jaime E. *Frenemies: How Social Media Polarizes America.* Cambridge University Press, 2019.

Siemaszko, Corky. "Trump's anti-media rhetoric turns to action". *NBC News.* February 12, 2025.

Simon, Joel and Robert Mahoney. *The Infodemic: How Censorship and Lies Made the World Sicker and Less Free.* New York: Columbia Global Reports, 2022.

Singer, Peter. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton: Princeton University Press, 2011.

Soave, Robby. *Tech Panic: Why We Shouldn't Fear Facebook and the Future*. New York: Threshold Editions, 2021.

Srivastava, Sarita. *"Are You Calling Me a Racist?": Why We Need to Stop Talking about Race and Start Making Real Antiracist Change*. New York: New York University Press, 2024.

Stanley, Jason. *Erasing History: How Fascists Rewrite the Past to Control the Future*. New York: Atria, 2024.

Stone, Daniel F. *Undue Hate: A Behavioral Economic Analysis of Hostile Polarization in US Politics and Beyond.* Cambridge: The MIT Press, 2023.

Sumner, L. W.. *The Hateful and the Obscene: Studies in the Limits of Free Expression*. Toronto: University of Toronto Press, 2004.

Sunstein, Cass R.. *Conformity: The Power of Social Influences*. New York: New York University Press, 2021a.

Sunstein, Cass R.. *Liars: Falsehoods and Free Speech in an Age of Deception*. New York: Oxford University Press, 2021b.

Sunstein, Cass R.. *On Liberalism: In Defense of Freedom*. Cambridge: The MIT Press, 2025.

Sunstein, Cass R.. *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press, 2018.

Szetela, Adam. *That Book Is Dangerous!: How Moral Panic, Social Media, and the Culture Wars Are Remaking Publishing*. Cambridge: The MIT Press, 2025.

Taibbi, Matt. *Hate Inc.: Why Today's Media Makes Us Despise One Another.* New York: OR Books, 2021.

Tosi, Justin and Brandon Warmke. *Grandstanding: The Use and Abuse of Moral Talk*. New York: Oxford University Press, 2020.

Turley, Jonathan. *The Indispensable Right: Free Speech in an Age of Rage*. New York: Simon & Schuster, 2025.

Ungar-Sargon, Batya. *Bad News: How Woke Media is Undermining Democracy*. New York: Encounter Books, 2021.

Vaidhyanathan, Siva. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. New York: Oxford University Press, 2021.

Vallier, Kevin. *Trust in a Polarized Age.* New York: Oxford University Press, 2020.

Vigdor, Neil, and Chris Cameron. "Nikki Haley Renews Call for TikTok Ban After Bin Laden Letter Circulates". *The New York Times*, November 17, 2023.

Walter, Barbara F. *How Civil Wars Start and How to Stop Them.* New York: Crown, 2022.

Whittington, Keith E.. *Speak Freely: Why Universities Must Defend Free Speech*. Princeton: Princeton University Press, 2019.

Willard, Mary Beth. *Why It's OK to Enjoy the Work of Immoral Artists*. New York: Routledge, 2021.

Williams, Thomas Chatterton. *Summer of Our Discontent: The Age of Certainty and the Demise of Discourse*. New York: Penguin Random House, 2025.

Zweig, David. *An Abundance of Caution: American Schools, the Virus, and a Story of Bad Decisions*. Cambridge: The MIT Press, 2025.