5-2022

Free Speech and Its Limits: An Exploration of Tolerance in the Digital Age

Jamie Forte
William & Mary

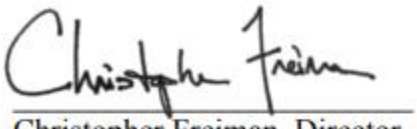
William & Mary Honors Thesis

Free Speech and Its Limits

An Exploration of Tolerance in the Digital Age

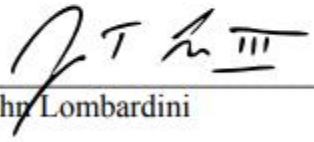
A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Arts from William & Mary by James Edward Forte

Accepted for Honors



Christopher Freiman, Director

Joshua Gert



John Lombardini

Williamsburg, VA
May 11, 2022

James Edward Forte
5-13-2022

James Edward Forte

Senior Undergraduate (2022) Majoring in Philosophy and Government

College of William & Mary – School of Liberal Arts

Advisor: Professor Christopher Freiman, Department of Philosophy

Thesis Title: Free Speech and Its Limits: An Exploration of Tolerance in the Digital Age

Thesis Abstract:

Humans have made remarkable strides in protecting and preserving free speech despite an overwhelming historical legacy of censorship and suppression of dissent. Given that history makes clear how easy it is to slide into authoritarianism and sacrifice our rights in the name of security, and given that we find ourselves frequently facing the temptation to do so, this is not an unreasonable position. If the United States is one of the few bastions of free speech in an otherwise unfree world, then we must defend this freedom vehemently, or so the argument goes.

While this position is not an unreasonable one, it diminishes much of the important complexity of the debate. For speech *necessarily can never be unlimited*, and speech can have serious consequences in many cases. With the advent of the internet and social media, now more than ever we must reconsider when and for what reasons we are unwilling to limit speech. If we cannot establish solid, consistent ways to make this determination, we will find ourselves paralyzed if and when limiting speech becomes necessary. In this paper, I will examine the various principles we might use to draw that line, their implications for speech, and especially how to apply such principles to the internet going forward, which presents today the greatest challenges to free speech that we face.

jeforte@wm.edu

Table of Contents

Introduction

Free speech occupies a place of central importance in modern liberal thought. It is considered an essential element of the open society, such that no society lacking it is held to be free. Yet, as with the debates over terms like "freedom" and "democracy", the meaning of "free speech" has become blurred over time. There are many in the United States, for instance, who believe that free speech is absolutely unlimited and there are no justifiable restraints on expression. Such views reveal a crucial lack of understanding as to what rights actually are, as rights in any society will inevitably come into conflict, and one right must be given priority over the other in that instance for the conflict to be resolved. At the same time, the internet has grown and evolved at a staggering rate, completely outpacing our ability to keep up regulatorily, much in the same way that big data has outpaced our privacy laws.[1]

The internet, which became functional with the adoption of TCP/IP in 1983 and publicly available in 1991, now plays an inextricable role in our public discourse, yet speech online is neither afforded the same protections nor subject to the same restrictions faced by speech in the physical world. Despite this, the internet is very much a part of the "real" world, and speech online can have very real consequences, whether in the form of disinformation or slander. This is further compounded by the fact that there is little to no accountability for websites' treatment of speech. These websites are, after all, mostly operated by private entities which have full discretion to run their sites as they please. Moreover, anonymity and the removal of barriers to publication undermine the premises upon which our truth- and consensus-formation

[1] Zuboff, *The Age of Surveillance Capitalism*.

mechanisms depend because they facilitate the spread of disinformation and enable the persistence of false consensuses. Unchecked this can lead to widespread disbelief in the legitimacy of government and electoral institutions and even result in insurrection or coup attempts, yet the government's principal disavowal of any orthodoxy or official truth leaves it ill-equipped to combat such movements until they reach the stage of action. This should very much worry us when we realize that our current "free marketplace of ideas" model is not very well equipped to deal with these problems either.

In this paper, I propose a *less* laissez-faire approach to the marketplace of ideas: increasing platform accountability for online speech through the reintroduction of liability for large platforms and the creation of additional regulatory constraints aimed at addressing problems like disinformation and false consensuses. I begin by looking at the history of free speech in western thought, starting with the Greeks and coming forward to the Renaissance and early modern period, before discussing the classic liberal positions of Locke and Mill. I then look at select legal cases in United States court precedent which were influential in the formation of our current interpretation of speech principles and how these cases were decided. Finally, I consider more recent trends that have developed with the internet and social media. These technologies pose challenges to open societies never-before experienced. I suggest some steps the U.S. might take to reduce tensions in public discourse without forcing Americans to surrender certain freedoms with which they are obviously loathe to part.

Chapter 1—A Brief Intellectual History of Free Speech

Any serious discussion of speech rights necessitates some consideration of their origins. We can certainly trace a loose lineage stretching back through 2500 years of intellectual history, but the exact nature of that evolutionary path and which ideas mattered most are subject to debate. There is general agreement that speech rights as we have them today first emerge in a somewhat recognizable form with the rise of liberalism. John Locke is often cited as a central figure for his presentation of social contract theory and emphasis on *individual* rights. It should be noted, however, that Locke—like many other political theorists—is engaging in a tradition stretching back to Plato.

Basically, it begins by asking a question: why do governments fail to live up to their normative obligations? The answer will obviously depend on the type of government in question, the circumstances of the particular case, etc. But the fundamental approach will always involve identifying which factors helped to make the particular regime successful, and which factors hindered its efforts. For much of history, thinkers have looked at speech in this manner. Does allowing some degree of speech freedom help the state maintain itself, or hinder it? Usually they concluded the latter. People could not be trusted with the freedom to speak their minds, because the exercise of such a freedom has the potential to upset the social structures which underwrite the society. How could the king and lords retain their dignity if any peasant could insult them? What would stop heretics from tarnishing the good word of God and spreading sinful lies? This freedom was much too powerful for the common man, and thus, as a source of danger to society, it had to be suppressed. The core of this argument is captured in Plato's *Republic* and *Laws*, when he talks about the dangers of subjective taste in the theater:

> "In those days Athenian music comprised various categories and forms…Once these categories and a number of others had been fixed, no one was allowed to pervert them by using one sort of tune in a composition belonging to another category…children and their attendants and the general public could always be disciplined and controlled by a stick. Such was the rigor with which the mass of the people was prepared to be controlled in the theatre, and to refrain from passing judgment by shouting. Later, as time went on, composers arose who started to set a fashion of breaking the rules and offending good taste. They did have a natural artistic talent, but they were ignorant of the correct and legitimate standards laid down by the Muse. Gripped by a frenzied and excessive lust for pleasure, they jumbled together laments and hymns, mixed paeans and dithyrambs, and even imitated pipe tunes on the lyre. The result was a total confusion of styles. Unintentionally, in their idiotic way, they misrepresented their art, claiming that in music there are no standards of right and wrong at all, but that the most 'correct' criterion is the pleasure of a man who enjoyed the performance, whether he is a good man or not. On these principles they based their compositions, and they accompanied them with propaganda to the same effect. Consequently they gave the ordinary man not only a taste for breaking the laws of music but the arrogance to set himself up as a capable judge. The audiences, once silent, began to use their tongues; they claimed to know what was good and bad in music, and instead of a 'musical meritocracy', a sort of vicious 'theatrocracy' arose. But if this democracy had been limited to gentlemen and had applied only to music, no great harm would have been done; in the event, however, music proved to be the starting point of everyone's conviction that he was an authority on everything, and of a general disregard for the law. Complete license was not far behind. The conviction that they *knew* made them unafraid, and assurance engendered effrontery. You see, a reckless lack of respect for one's betters is effrontery of peculiar viciousness, which springs from a freedom from inhibitions that has gone much too far…This freedom will then take other forms. First people grow unwilling to submit to the authorities, then they refuse to obey the admonitions of their fathers and mothers and elders. As they hurtle along towards the end of this primrose path, they try to escape the authority of the laws, and the very end of the road comes when they cease to care about oaths and promises and religion in general." (*Laws* 700b-701c)[2]

Plato's underlying concern here is that increasing subjectification of taste leads to the subjectification of societal norms, which will ultimately result in a breakdown of all values, yielding a society without any sense of cohesion or regard for the common good. While he is unwarranted in claiming that the connection is *necessary*, it is nonetheless a possible result of rampant, unchecked individualism. I have quoted this passage here at length because it is a model for most conservative arguments from potential consequence and captures the main

[2] Plato, *Complete Works*, 1389–90.

reasons speech rights were feared and shunned: they undermine order and belief in shared values; they (like other individual freedoms) encourage citizens to look inwards, rather than outwards to society; and perhaps worst of all, they give citizens the license to think themselves capable arbiters. This last point was of particular concern for Plato, as he was very much worried about people who are unaware of their own ignorance. Thus, we get many concerns which stem from the ignorance of the public: citizens do not *really* know what is best for themselves and for society, and we shouldn't let them spread the wrong ideas; the common people are too easily manipulated and unchecked speech presents the risk that villainous men will abuse it; etc. the list goes on. Perhaps the only reason speech rights arguments managed to gain some ground was because of the emphasis liberalism places on individuality, and the framework it provides for arguing for its importance. The point I am trying to illustrate here is that "rights" themselves presuppose some principled regard for the individual. You will be hard-pressed to find them in strongly collectivist or authoritarian societies.

The Sophists, Plato's longtime opponents and the first Panhellenists, held much more open and democratically friendly positions, especially regarding notions of truth, political expertise, and who should make important decisions. As we can see in the example of Thrasymachus in the *Republic*, the sophists were more in line with the contemporary Athenian view that some things, like political expertise, were universal such that any man could legitimately weigh in on those matters.[3] Plato naturally disagreed, asserting that since there is an objective truth which can be known through reason, there is naturally also a correct answer even in matters of political expertise, such that we are mistaken when we allow just anyone to guide

[3] Sørensen, *Plato on Democracy and Political Technē*, 12–18, 28.

us. While liberalism places us more in line with the sophistic view, the historical prevalence and practical appeal of Plato's approach is worth keeping in mind.

It is important that we not conflate speech rights with tolerance. The former are a somewhat unique product of liberalism, whereas the latter can be found in all sorts of regimes and societies. Ancient Rome famously practiced a kind of tolerance, wherein citizens were allowed to hold whatever views they desired and express themselves, provided they accepted the authority of the Empire and accorded with public and ritual practices. Christians earned the ire of Roman authorities by refusing to participate in official state rituals, a political act tantamount to rebellion against Empire, which led to their persecution. They were not tolerated because they refused to accord with Roman practices, unlike the Jews. In democratic Athens, citizens had the right to participate in politics, but nothing in the way of protections against the exercise of state power. As in the trial and death of Socrates, citizens who had earned the condemnation of the crowd could expect little sympathy or support. Citizens of the empire under the reign of Marcus Aurelius could criticize the *princeps*, but only under the auspices of the Emperor's tolerance and goodwill. There was nothing stopping him from changing his mind and enforcing censorship at his discretion.

Speech rights introduce a very interesting tension into political governance, insofar as liberal societies deem them to be necessary despite the many risks which accompany them. Let us consider those risks (and advantages) from a non-normative standpoint. Spinoza and Mill both present arguments which assert that suppressing dissenting speech is ultimately ineffective and unsustainable. Mill's is the more famous position, often referred to as the "free marketplace of ideas". I intend to discuss Mill more thoroughly later in this chapter, but his principal argument is that bad ideas cannot be effectively defeated unless they are allowed to be openly heard and

challenged. Force is not an argument, and as such cannot win a debate. Likewise, the only way to definitively quash an argument is to show how it is flawed. By silencing the other speaker, you might temporarily suppress his views, but you haven't demonstrated *why* they are wrong. Mill thinks that by allowing all opinions to be aired in a public space (*especially* controversial views), natural competition between ideas through their proponents and detractors will weed out the worst ideas until only the best (or most worthy) remain, as bad ideas will be challenged and subsequently defeated. Naturally, whether or not this actually occurs will depend on the society in question and quality/nature of its discourse. So to assert that it will hold true requires an assumption of faith on the part of the speaker—a conviction in the ability of the wisdom of the crowd to triumph over ignorance.

Spinoza is far less optimistic. He believes that free speech is necessary for a similar reason to Mill, in that a regime which responds to legitimate verbal challenges with arbitrary force undermines its own legitimacy, since the regime's seeming inability to argue back coupled with its determination to crush the opposing perspective instead indicates that said perspective may have some merit to it. In other words, regimes must be able to answer ideological and philosophical challenges with reason as well as force, because legitimacy depends on people's perception. But in accordance with the paradox of tolerance, an unlimited degree of speech freedom presents opportunities for factions looking to undermine the regime and seize power, and there is no need for those factions to "play by the rules" or present valid arguments, provided they have some degree of popular support. For this reason Spinoza saw fit to appoint a sort of established orthodoxy to defend the state from existentially dangerous ideological threats.[4]

[4] Spinoza, *Theological-Political Treatise*, 257–59.

This concern—that free discourse would undermine regime power and create avenues for counter-regime propaganda—pops up majorly around the invention of the printing press. If anyone can publish anything, what is to stop the common people from being fed lies? The new printing press with moveable type was a powerful new tool, and a powerful threat in the wrong hands. Most of these concerns ultimately proved relatively minor, although the long list of "pamphlet wars" speaks to the power free printing has to distribute and spread ideas. Long before the rise of classical western liberalism, the Catholic Church dominated European thought and exerted absolute control over the truth. Numerous works deemed dangerous or heretical were banned. Starting in the early 16th century Catholic city-states started publishing early versions of the *Index Librorum Prohibitorum*, or "Index of Prohibited Books". First in the Netherlands in 1529, then in Venice in 1543 and Paris in 1551, the *Index* soon spread throughout Europe and came to symbolize the Church's determination to control public discourse and quash any heresy before it could gain traction. Even prior to the *Index*, Christian theologians argued extensively over which pagan works should be censored or allowed, with a common point of conflict being whether to allow the transcription and distribution of classical works like the philosophy of Plato. Platonic and Aristotelian arguments often meshed well with the Christian tradition, and so the works often caused divisions over whether they should be banned or allowed. Some, like St. Basil (c. 330-379) argued against the censorship of pagan classics on the grounds that they promoted virtue as well as vice, and that it was the duty of the reader to take away only the good parts from the text. Others, like Giovanni Dominici (1356-1420) saw no value in anything outside of scripture and attempted to institute a blanket ban on pagan literature.[5]

[5] Grendler, *The Roman Inquisition and the Venetian Press*, 1540-1605, 65–66.

The rise of Protestantism played a huge role in breaking the Church's monopoly over truth. The rediscovery of previously lost copies of important texts prompted a renewal of interest in textual scholarship and the origins of the law, resulting in competing interpretations and intense debates over the meanings of texts and the original nature of the primitive Christian church. Fierce conflicts arose over all points of scripture, and the fighting was no less fierce over trivial matters than it was over matters of salvation. The tiring nature of such constant conflict in turn prompted second-level consideration of how to conduct such debates, to what extent the Church could tolerate disagreement over the interpretation of scripture, and if so which matters required concordance and which allowed for dissent. Erasmus advocated for more religious toleration, noting that overly fierce debates inhibit their participants' ability to arrive at the truth. Legal theorists like Grotius and Pufendorf similarly found themselves embroiled in such debates, and reflection thereon gave rise to more formal doctrines of religious toleration. At the same time, secularism was on the rise in Europe, with more and more people starting to look to reason and empirical studies of the natural world for guidance, rather than religious authorities.

These trends helped to spur publication of books in the vernacular. While there were still many checks to publication, including editors, publishers, financial barriers, and censorship by authorities, there is no doubt that the printing press greatly democratized publishing power, and radically transformed how information was disseminated in Europe. By the end of the 16th century, however, we can also see widespread concerns regarding how printing power could be used to undermine state and religious authority. The English Licensing of the Press Act of 1662 and the various legislative battles for control of publishing power that preceded it are good examples of this concern. Two years prior to the start of the English civil war, the Habeas Corpus Act of 1640 abolished the Star Chamber and its 1637 Decree, which for our purposes had

the effect of removing most prepublication censorship in England. Three years later, however, parliament passed the Licensing Order of 1643, reinstating all of that censorship, just under Parliament's terms and direction rather than the Star Chamber's. John Milton (1608-1674) published his *Areopagitica* the following year in response, arguing against the Licensing Order and in particular against pre-publication censorship. While Milton was himself no advocate of general free speech, many of his ideas regarding the importance of open public discourse and the open vetting of disagreement continue to hold sway, and are present in Mill's position.

Milton begins by justifying the need for constructive criticism (which he contrasts with mere flattery), and cites both biblical examples and recent history (like the Inquisition) to argue that reading bad or heretical works is a necessary part of a proper education, and that pre-publication censorship is easily abused by authorities for political purposes, as can be seen from many examples of Catholic Church behavior. Most importantly, perhaps, Milton argues that people are not necessarily corrupted through exposure to bad ideas, which can even be an impetus for better ones, and thus judgment and acceptance of a particular view should always be left to the discretion of the reader, not decided on his behalf by the authorities. On more pragmatic grounds, Milton argues that the Licensing Order is incapable of actually preventing societal corruption (as there are many other channels through which ideas can spread such as word of mouth, and most common people don't read books anyway), and in a manner similar to the slippery slope, Milton sees no terminus point for additional measures to be implemented in the name of furthering this goal (i.e. reducing societal corruption). Milton directly challenges Plato's regime of censorship as laid out in the *Laws* passage, arguing that only proper education can rectify societal corruption:

> "Nor is it *Plato's* licencing of books will doe this, which necessarily pulls along with it so many other kinds of licencing, as will make us all both ridiculous and weary, and yet frustrat; but those unwritt'n, or at least

> unconstraining laws of vertuous education, religious and civill nurture, which *Plato* there mentions, as the bonds and ligaments of the Commonwealth, the pillars and the sustainers of every writt'n Statute; these they be which will bear chief sway in such matters as these, when all licencing will be easily eluded." – Milton, *Areopagitica*[6]

Asserting that he "cannot praise a fugitive and cloister'd virtue, unexercised & unbreath'd," Milton goes on to argue that there is no glory in unscrupulous victory in matters of truth:

> "When a man hath bin labouring the hardest labour in the deep mines of knowledge, hath furnisht out his findings in all their equipage, drawn forth his reasons as it were a battell raung'd, scatter'd and defeated all objections in his way, calls out his adversary into the plain, offers him the advantage of wind and sun, if he please; only that he may try the matter by dint of argument, for his opponents then to sculk, to lay ambushments, to keep a narrow bridge of licencing where the challenger should passe, though it be valour anough in souldiership, is but weaknes and cowardice in the wars of Truth. For who knows not that Truth is strong next to the Almighty; she needs no policies, nor stratagems, nor licencings to make her victorious, those are the shifts and the defences that error uses against her power: give her but room, & do not bind her when she sleeps, for then she speaks not true, as the old *Proteus* did, who spake oracles only when he was caught & bound, but then rather she turns herself into all shapes, except her own, and perhaps tunes her voice according to the time, as *Micaiah* did before *Ahab*, untill she be adjur'd into her own likenes." – Milton, *Areopagitica*[7]

For Milton, truth does not need coddling, and truly strong positions welcome challengers. If one is confident in the truthfulness of his position, he should have no reason to fear criticism or objections to it, as his truth will be demonstrated eventually. Likewise, if one really cares about the truth, he has every reason to hear out legitimate challenges, through discussion of which we may further refine our truth. Resorting to censorship is thus a kind of intellectual cowardice, for those who genuinely pursue the truth have no reason to fear debate.

It is with this backdrop that we consider John Locke (1632-1704). With his theories of natural law, the social contract, and his hopeful (at least compared to Hobbes') presentation of the state of nature, Locke sketched out the foundations of liberalism. Through a natural system of

[6] Milton, *Areopagitica*.
[7] Ibid

obligations Locke shows how men can live harmoniously without eschewing most of their freedoms. Underpinning Locke's arguments are his notions of the state of nature and the natural law which governs it. Natural law is "natural" insofar as it can be logically deduced from first premises and thus is always in effect wherever there are humans; for this reason it constitutes a higher basis for authority than any person or institution. Locke believes natural law is descriptive of human nature but lacks prescriptive force, as men may still choose to violate it at risk of being punished by their fellows. Hence, while men may be free in the state of nature, and may even abide by natural law, they also live with the constant fear that an appeal to natural law will not save them from the most determined wrongdoers—hence the need for civil society to protect their persons and property.[8]

Having already dismantled Filmer's *Patriarchia* in his *First Treatise*, Locke argues that natural law provides the true basis for the state's legitimacy: government is only legitimate when it has the consent of those whom it governs, through the mechanism of the social contract. Government retains its legitimacy by continuing to serve and represent the people's interests, and—most crucially—if at any point in time the government violates this contract, any obligation on the governed to obey its edicts is rendered void. All people thus have a natural right to revolution, provided that they are being unduly restrained by a tyrannical and arbitrary power. Locke thus grounds the basis for governmental authority on the consent of the governed and presents a view of government as existing only to facilitate civil society and preserve men's freedoms, which are present but insecure in the state of nature. Naturally, rights play a central role in his framework, though Locke is primarily focused on property rights and does not extend all the same considerations to speech. His arguments in support of religious toleration can

[8] Locke, *Second Treatise Of Government*.

nevertheless be extended to support toleration of heterodoxy more generally, and his emphasis on limited government lends itself to principled non-interference in people's thoughts and lives wherever possible.

In making these arguments in his *Letter Concerning Toleration*, Locke draws a subtle distinction between religious beliefs and political principles attached to them. This allows him to condemn Catholic political practices (like subservience to the Pope) without necessarily denying Catholic worship more broadly. This provides much of the basis for separation of Church and state, as it follows that nearly any religious belief or practice can be allowed provided it is not inimical to toleration. In this regard, toleration itself becomes a sort of criterion for the limits of dissent, as without toleration there can be no civil society. Locke does not extend this toleration to atheists, but his concern over the indeterminacy of religious truth supports some individual choice between interpretations. Most crucially for our purposes, as every man is his own judge and interpreter of natural law in the state of nature, it follows in Locke's framework that the individual can be a capable judge and thus has the right to think and decide for himself wherever and whenever possible.

The French philosopher Voltaire (1694-1778) is worth mentioning here because he often pushed the absolute limits of tolerance with his scathing satires, which critiqued French customs, dogma, and figures alike. Voltaire frequently found himself in trouble for making fun of authority figures: he was imprisoned in the Bastille three times, exiled to England for a period of two years, and forced to flee or outright banned from Paris on at least three separate occasions. This only fueled his resentment of fanaticism, dogma, and the arbitrary exercise of power. After a close reading of the bible in which Voltaire found its authority dubious at best, he formulated a strong stance on the need for religious toleration, a position he would hold for the rest of his life.

Voltaire's many parodies are poignant reminders of the genre's power to convey important social and societal critiques, and how libel laws can be easily abused to suppress legitimate dissent in the absence of proper rule of law.

John Stuart Mill (1806-1873) offers the strongest defense of free speech as a principle, and perhaps goes further than anyone else in emphasizing the importance of defending even and especially the most controversial speech. Mill's presentation of and justifications for free speech come closest to embodying our modern notions. He is antithetical to Plato, insofar as he begins with an assumption Plato outright rejects and fears: the presumption that individual autonomy is important and that individuals can be capable judges of their own well-being. This presumption is implicit in Milton, Locke, and Spinoza, but Mill asserts it in no uncertain terms. Mill asserts the classical libertarian position that paternalistic restraints on individual liberty (i.e. ones which presuppose that the agent is incapable of judging what is best for himself) must be heavily justified and subject to a presumption of restraint. In his words, "If all mankind were of one opinion, and only one person were of the contrary opinion, mankind would be no more justified in silencing that one person, than he, if he had the power, would be justified in silencing mankind."[9] No degree of consensus is sufficient to justify the suppression of unpopular or even loathsome ideas. Mill's reasons for this are many, but chief among them is his concern that the suppression of ideas will result in a stolid and stagnant public discourse—one in which complacency and intellectual laziness prevent us from adequately questioning our own knowledge, such that we close ourselves off from the truth. The worst thing that can happen, Mill thinks, is a world in which we reject Galileo and Copernican heliocentrism simply because

[9] Mill, *On Liberty*, 18.

we are more comfortable with Aristotelian geocentrism. A strong and truth-regarding public discourse is thus essential to the continued advancement of human knowledge.

Mill does of course consider limit cases for the general presumption against restraints upon individual liberties, and in doing so formulates his "harm principle," which stipulates that the liberties (such as speech) may only be restrained when they result in instances of actual harm or otherwise infringe upon the liberties of others. While Mill is not entirely clear on how to define harm (indeed, much of the debate on this subjects consists therein), he is nevertheless adamant that the bar for harm must be set fairly high. In the example of the corn dealers, Mill permits speech which indirectly disparages or harms corn dealers, but does not permit speech which directly results in the corn dealers being harmed, such as telling an angry mob that corn dealers are trying to starve them. This example is particularly salient when we consider that it appears to capture even the imminent lawless action aspect of the standard set forth in *Brandenburg v. Ohio* (1969), which I will discuss in chapter 2.

Let us consider Mill's working assumptions here. The root of Mill's advocacy is his conviction that public discourse is necessary to refine our understanding of the truth, and that unjustified restraints on individual liberty are morally impermissible. Let us assume for now that both of these assumptions hold true. The problem lies mainly with Mill's third assumption—that a sufficient proportion of the population uses reason to assess the truthfulness of statements and ideas. We might refer to this as "scientific thinking", which is to say the idea that one's support for a position is contingent upon the truthfulness of the assumptions underpinning that position. I might support a tax on carbon emissions, for example, on the condition that it is an effective strategy for reducing emissions. In this instance I care primarily about the reduction of emissions, and, if I am presented with information suggesting that carbon taxes are ineffective at

reducing emissions, I will cease to support a carbon tax, as my support for that policy is contingent on its ability to fulfill my actual objective of reducing carbon emissions. In other words, when presented with information which indicates that my position is inconsistent and/or incorrect, my response should be to update my views on the basis of new information. This willingness to revise my views in the face of being proven wrong is absolutely essential to my ability to arrive at a more correct conclusion. Accordingly, if I am unable to do this (assuming that I have been provided with sufficient evidence to invalidate my prior perspective) I will be unable to revise my position to a more correct one. Of course this may not matter in the macro scale if enough people *do* correctly revise their views, but history, recent events, and research in evolutionary psychology all suggest that humans are not adapted to prioritize truth over other factors, like interpersonal relationships.[10] Moreover, if the reasons a person holds for supporting a given position are not grounded in evidence but rather some other source of conviction (i.e. their support for a position is not contingent on said position's truthfulness or accordance with some other doctrine), even when their position is factually invalidated they will not revise their views. This phenomenon largely explains the prevalence of disinformation today and so-called "zombie statistics," and also explains how some views, like Naziism, which are antithetical to the society's values and have already been summarily defeated in public debate, nevertheless continue to hold sway and attract new followers. If demonstrating the factual inaccuracy of white supremacy was enough to turn away its adherents, it would not continue to rear its ugly head. Such examples should make us hesitant to agree with Mill that debate will always result in a refinement of positions and ideas in the aggregate. Clearly, there is a whole host of other factors

[10] Kolbert, "Why Facts Don't Change Our Minds."

which influence public discourse in ways which are difficult to measure, let alone predict, and we would be amiss without considering these in more detail.

It is interesting to note that many of the arguments employed in these debates can be similarly wielded or repurposed for completely different causes. It is worth expanding upon here because we can see it give rise to a few different general "types" of pro- or anti- free speech arguments. Even without assuming a tendency toward more liberty, all anti-speech arguments take the form of an argument from concern. For example, "We should ban this book because it tells dangerous lies to the people."

The concern here is that the population will start to believe "untrue" ideas, which is itself more or less of a concern depending on what those ideas are. Authorities are likely to be more concerned by a pamphlet like Paine's *Common Sense* advocating the overthrow of non-representative governments than they are by a flyer claiming that the moon is made of cheese. The counterpart to this concern is that restricting access to books based on content is a paternalistic overreach by the state, and that it is wrong to presume that the state knows truth better than the common man. The more philosophically-inclined may even argue that truth is subjective, or at the very least not known to any one man or group in its entirety, and that it is therefore unwarranted for the state to claim a monopoly over its definition and interpretations.

Another line of argument often seen stems back to Plato, and it is helpful first to explain this aspect of his philosophy. Plato believes that there is an objective, eternal mind-independent Truth which can be known through reason. It follows from this that there is a "right" answer to questions in areas such as morality, with all other answers being wrong. This goes a long way in explaining Plato's otherwise strange notions of freedom, including why he thinks that men are only free when they live their lives in accordance with reason. In very simple terms, any choice

not informed by this rational truth becomes a folly of ignorance, and the argument to limit people's speech is justified on the basis that they don't know any better or are otherwise slaves to their own ignorance. Spinoza strongly echoes this Platonic line of reasoning, positing that men are only free when they live their lives in accordance with reason, and that therefore nobody should want to do anything reason condemns. In other words, freedom is equated with a kind of obedience to one's higher consciousness, wherein individuals become free by forming adequate ideas about themselves and do what is best out of an enlightened understanding of their own interest. To live wantonly or ignore "what is best" would be to place oneself in a lower state of freedom, since one is a slave to ignorance.

This logic is easily extensible and can be used to justify suppression of any speech which contradicts an established dogmatic "Truth". Concern (that censorship will be employed to suppress what nevertheless needs to be heard) is far more common than its counterpart, so I will provide an example of the latter here for clarity. While in western cultures we have generally come to accept that no one has a monopoly over truth, we nevertheless consider scientific consensus to be objective truth, and (typically) treat it with the same necessity and forcefulness that we would a posit claiming metaphysical or theological certainty.

I consider this tension between the power of the collective to assert an orthodoxy and the power of the individual to reject that orthodoxy to be a key component in the debate over free speech. However we may feel about Plato, we are nevertheless obligated to care about what is true, because the truth has real and practical consequences. At the same time, history gives us myriad examples of that power of orthodoxy being abused to crush dissent. Had the Catholic church prevailed against Gallileo, we might very well still believe that the Sun orbits the Earth, simply because that is what Ptolemy says (even though it is demonstrably false).

By now, we should have a rough understanding of how speech rights evolved and the important doctrines which underpin them. At risk of oversimplifying, much of the extant debate over free speech can be reduced to a measure of the strength of one's conviction in the wisdom of the crowd. Some thinkers, like Plato, were vehemently pessimistic, and saw only chaos following the wanton exercise of individual liberty. Others like Milton, Locke, and Mill believed in the collective's power of judgment, and saw individual autonomy as being worth the risks that accompany it. Yet others still, like Spinoza, saw individual autonomy as important but lacked faith in society's ability to overcome superstition and prejudice. On this particular point there is no definitive answer, as the question is empirical by nature, and only further history can resolve it. Nevertheless, the terms of the debate should be clearer now, as should be the basis for justifying speech rights.

I would like to conclude this chapter by skipping ahead to the present and offering three modern philosophical principles used for determining the limits of speech, as these will become relevant in chapter 2. The first principle is Mill's harm principle: speech rights can be limited only when their exercise causes demonstrable harm to others. This is the most widely instantiated principle for limiting speech and is also the main basis for limiting speech in the United States. While there is much room for debate over what counts as harm, that harm should be the boundary for speech is relatively uncontroversial.

The other two principles address cases not covered by the harm principle, and thus should be taken as supplementary. The second is Feinberg's offense principle, which justifies limiting speech on the basis of offense it causes to others, with additional considerations for context and circumstances. In his book, *Offense to Others, The Moral Limits of Criminal Law*, Feinberg gives the following definition: "It is always a good reason in support of a proposed criminal

prohibition that it is probably necessary to prevent serious offense to persons other than the actor and would probably be an effective means to that end if enacted."[11] The offense principle is more controversial than the harm principle, as the bar for offense is much lower than for harm, especially when considering that it is often the case that at least one person will take offense over anything. For this reason, Feinberg stipulates that penalties for offense must always be lower than they are for harm, and indicates that it is necessary to consider a whole host of other contextual factors in each case, such as the "extent, duration and social value of the speech, the ease with which it can be avoided, the motives of the speaker, the number of people offended, the intensity of the offense, and the general interest of the community."[12] For instance, in cases where the offense is in a form which can be easily avoided, like a book, there is little justification for banning it, whereas in the case of a pornographic billboard, the offense cannot be avoided and thus the community has a legitimate basis for restricting or preventing its placement. No one is forced to read a book they find offensive; conversely it is nigh impossible to avoid seeing an offensive billboard. Other factors like social value help to capture distinctions between speech which may nevertheless be of interest to us. For instance, we generally consider debate and discussion to have a higher social value than a shouting match, as the former involve an exchange of ideas while the latter produces nothing of value. While there is no doubt difficulty in determining the severity of an offense, or what counts as a sufficient number of offended people to warrant further action, we nevertheless already practice some form of the offense principle with public indecency laws and prohibitions on lascivious advertising.

[11] Feinberg, *The Moral Limits of Criminal Law*, 2:xiii.
[12] van Mill, "Freedom of Speech."

The third principle is the one articulated by Stanley Fish in his book *There is No Such Thing as Free Speech*, which for the purposes of this paper I will call the "democratic preference principle". Borrowing Fish's paraphrasing of a line from Canada's constitution, we may define this principle as follows: "every right and freedom herein granted can be trumped if its exercise is found to be in conflict with the principles which underwrite the society."[13] In essence, this principle asserts that speech which somehow violates or undermines the society's foundational principles (i.e. equality under the law, commitment to electoral institutions, etc.) may be justifiably limited, as it has the potential to undermine the principles which make that society possible.

This principle is naturally the most controversial of the three, as much depends on both our definitions of our foundational principles and on how we define speech that contradicts them. We can, however, bolster it by adding similar considerations to those Feinberg makes use of in her offense principle. For instance, the size of a speaker's audience—we are only concerned about threats in the abstract when they attract a substantial following and thus signal more concrete action. This principle covers both speech which actively foments rebellion and speech which violates principles in the abstract in such a way as to undermine the very procedural mechanisms of the society, in effect precluding further discourse.

To illustrate this point, consider the various (sometimes implicit) agreements which underpin the functionality of our society. Toleration of dissent, for instance, is crucial, as without it there can be no open or free discussion. Likewise, the agreement to resolve disputes through legal means and third-party arbitration rather than resorting to force renders any extralegal action to this effect illegitimate. These kinds of agreements form the boundaries of what kind of

[13] Fish, *There's No Such Thing as Free Speech, and It's a Good Thing, Too*, 105.

behavior can be tolerated. Conversely, stepping outside these boundaries effectively undermines the whole process, as the process is only fair when everyone abides by these restraints upon their action. Governments are held to be legitimate when elected because of a shared agreement to make peace with the election's result, however personally unfavorable it may be; without this agreement, the basis for legitimacy dissolves.

The bar that speech should meet to justify the use of this principle must accordingly be set very high. Any definition of the limits of this principle must be based on textual/legal principles with a strong consensus of interpretation, so as to avoid placing the total power to decide what counts as undermining in the hands of a few. The main concern with this principle is thus that we would need to limit its application so narrowly as to heavily overlap with or be completely covered by already-present speech cases like intimidation, threat, or incitation of violence, in which case the usefulness of instituting such a principle becomes suspect. Like the debate over whether there is a need for hate speech legislation, this question is not easily resolved. Nevertheless, the legal cases I examine in chapter two will help to illuminate the answer.

Chapter 2 – The American Legal Perspective on Speech and the Court Cases Which Defined It

Liberal democracies all have some legal protection for free speech, though varying in form and sometimes unwritten. There is a consensus that free expression has necessary limits, such as in cases of national security or imminent threat of harm, but countries disagree about what those limits are and which cases warrant the overriding of free speech. The United States in accordance with the First Amendment has the strongest pro-expression stance, allowing free speech to be suppressed only in cases where there is a clear and demonstrable imminent threat of harm such that there is not time to afford the public vetting of the information in the free marketplace of ideas. An application of Mill's harm principle, this stance is principally content-neutral: it forbids discrimination or suppression of speech on the basis of its content, with the exception of cases which fail the harm principle. In other words, all speech in the eyes of the law must be treated equally, however loathsome or offensive. Defenders of this position often make an appeal to Mill's free marketplace of ideas, arguing that it is not the law's role to judge and vet ideas—that role belongs to the realm of public discourse. Implicit in this argument is an assumption of faith in the power of the public ultimately to pick out the good ideas from the bad ones. Radically offensive speech thus is not a threat because the free marketplace of ideas will ultimately condemn such speech, and attempts to ban it legally only preclude further discussion.

I believe this view underestimates the power of disinformation. Invoking Karl Popper's Paradox of Tolerance, there is a necessary limit to what kind of speech can be tolerated, because speech which successfully undermines the foundational institutions underpinning public discourse effectively renders the process of debate moot, returning us to the state of nature. In

other words, shared agreement in electoral institutions and processes, refraining from violence, and practicing basic rules of civility at some minimum are necessary to preserve an open society; therefore there must be a limit, however generous it may be. We may find that we can afford to tolerate nearly every kind of speech, but when an idea gains force so as to become a powerful threat, principled, universal tolerance can hinder us from taking the steps we need to target and eliminate that threat. This extreme promise of equality under the law renders our legal tools blunt and imprecise, and we are paralyzed with inaction out of fear that any action we take will somehow reduce freedom elsewhere.

It is not surprising that the United States often struggles to produce clear and satisfactory speech laws. Other nations do not share its content-neutrality with regard to speech. Germany and France explicitly ban the sale and public display of Nazi paraphernalia. India has several overlapping laws which allow the government to suppress speech out of drastic concern for public order. Canada has a clause in its constitution allowing the government to trounce freedoms when their use is in conflict with the principles that underwrite Canadian society, and as of March 2022 has banned holocaust denial. These countries have not solved all the problems of free expression, and the difficulty of defining problematic speech and the need to consider context remain large challenges. Nevertheless, these nations have a great deal more latitude when it comes to dealing with problematic speech since they are allowed to single out troublemakers under certain conditions, and account for extenuating circumstances or the pervasive presence of historical legacies (such as slavery or Naziism).

We in the United States cannot do this. While the First Amendment provides the greatest protection of free speech, it also severely limits our ability to fight disinformation and assert the truth of any situation. This is not to say that we welcome subjectivity. Empirical science, for

instance, we take as truth. We also hold certain social and moral values as indisputable, such as prohibitions on child abuse. Even though it seems to be a massive breach of freedom to separate a child from its parents, in cases where the parents are unable to care properly for the child, the state nevertheless intervenes. In other words, while we may not ever be completely certain about the truth, we nevertheless have things we are very sure of, and being wrong about those things can have very real consequences. In the case of child custody, we place greater value in protecting children from harm than we do in respecting the parents' difference of opinion regarding what counts as abuse. While it may seem paternalistic or contrary to freedom, that only holds true within an entirely negative, purely theoretical view equating freedom with the power to act wantonly. In reality, we consider freedom to also include protection from certain negative externalities, like physical harm. We cannot dismiss Plato for thinking that people who harm themselves in their ignorance are less free, even when they do so of their own free will.

Like the debate over hate speech, free speech legislation is plagued by a series of definitional quagmires. How do we define which speech is protected, and which speech is not? How do we word restrictions so they can't be abused to silence protected speech later on? How do we make sure that restrictions meant to improve freedom in one area don't inadvertently reduce freedom in another? There isn't a clean and simple answer to any of these questions, which is what makes codifying speech law so difficult. Then there is the issue of vagueness. If a law is too specific, it won't apply in most cases, and an impractical and inordinate number of highly specific laws would need to be created to capture every scenario; those laws would need to be formulated and applied in a way that is consistent throughout. Conversely, if a law is too vague, it won't provide much guidance on what to do in a particular case, and thus be of little use; even worse, it might apply overly broadly and thus have potential for abuse. It is extremely

difficult to achieve the right balance between being specific enough to offer guidance on how to assess each case and being vague enough as to apply universally. Perhaps this is due to the number of variables involved: context is so important to determining speech cases that it is very difficult to generalize anything but the vaguest principles concerning speech without somehow unduly restricting it. The vagueness doctrine addresses this problem but does not solve it. Instead, it simply requires that for a law to qualify as constitutional, it must 1) state explicitly what it mandates and what is enforceable, and 2) define potentially vague terms. This effectively makes any overly vague law unconstitutional, though the courts still must first decide whether the law in question is sufficiently vague.

Despite the First Amendment's strict and strong defense of speech, we nevertheless limit it frequently out of practical necessity. We can see this manifested in restrictions on offensive signage and advertising, and public indecency, among other things. But we nevertheless take the right of protest very seriously. In the famous *National Socialist Party of America v. Skokie,* 434 U. S. 1327 (1977) case, for instance, the Nazis were ultimately granted the right to march in Skokie, Illinois, and only opted not to march out of a concern for the potential violence that could ensue when 12,000 angry counter-protestors showed up (it is also possible that the Nazis were trying to use the Skokie verdict to establish permission to march in Chicago). The court acknowledged that the Nazi march was incredibly offensive to Skokie's population, especially in light of it being more than 50% Jewish and having 5,000+ holocaust survivors, but nevertheless granted the Nazis the right to march, because the court concluded that offense was an insufficient criterion and swastikas fell short of the definition of fighting words established in *Cohen v. California,* 403 U. S. 15 (1971). Precedent regarding free speech law has been shaped by a number of such cases, not all of which resulted in the derivation of some speech-regarding

legal principle. Oftentimes it is easier to avoid ruling and thus setting a precedent by sidestepping the issue altogether. Just as human development is influenced by both genetic and epigenetic factors, the substance of free speech law is shaped not only by what the laws say but how they are interpreted and applied. It is therefore necessary to consider the precedent set by previous speech cases, as well as the nature in which each case was resolved (or in some cases, the issue sidestepped).

Before we begin, it is worth explaining something called *strict scrutiny*. When judges review a law which is in potential violation of a constitutional statute, they apply the strict scrutiny test, and if the law passes the test, it is found to be constitutional. Most laws fail this test, however, as it sets a very high bar. To pass strict scrutiny, a law must be narrowly tailored to achieve a compelling state interest, and it must also be the least restrictive method of achieving that interest. The first part ensures that any law challenging constitutional freedoms must have a strong and justifiable purpose, and the second part ensures that any restriction on freedom a law makes is necessary and not otherwise mitigatable or avoidable. As all of the cases I examine here involve some conflict with the First Amendment, it is important to keep the strict scrutiny standard in mind when considering how these cases were resolved.

In *Schenck v. United States,* 249 U. S. 47 (1919), Oliver Wendell Holmes first articulated his version of Mill's harm principle, which would serve as the guiding principle and litmus test in speech cases to come. Charles Schenck and Elizabeth Bauer were socialists who distributed leaflets urging citizens to disobey the draft during World War I, arguing that forced conscription violated the Thirteenth Amendment's prohibition on involuntary servitude. The government arrested Schenck and Bauer on charges of conspiracy to violate the Espionage Act of 1917, on the grounds that they were attempting to disrupt and undermine recruitment for the war effort.

Schenck and Bauer appealed, but the Supreme Court ultimately sided against them in a unanimous decision, in which Justice Oliver Wendell Holmes likened distributing the leaflets to the act of falsely shouting "fire" in a crowded theater. This created the notion of a "clear and present danger test"—the idea that the government can justifiably limit speech when its exercise creates a clear (identifiable and concrete) and present (imminent and certain) danger, like shouting "fire" in a crowded theater with the obvious result of the people inside panicking to get out and likely injuring each other in the process. To give another example of speech which would fail this test, inciting an angry mob to take action using unmistakable language would constitute a clear and present danger—namely that the mob will attack. Of course, we may wonder about the exact definitions of "clear" and "present". In Justice Holmes' example, the danger is clear (i.e. a stampede in the theater), and present in that it will occur immediately after "fire" is shouted. But in *Schenck v. United States*, the case this principle was first tested on, the clarity and imminence of the danger are less clear. There was no threat of immediate physical harm, because distributing pamphlets was merely a method of persuasion, in this case, persuading people to not support the war effort. There was no risk of physical harm as in the theater (unless you think the pamphlets alone could cause the United States to lose WWI), nor was the threat immediate (people have to read *and* be persuaded by the pamphlets before they start to have any effect).

Some will argue here that the actual harm in the theater example is caused by the stampede itself, rather than shouting fire, as people must hear the utterance and react to it for a stampede to occur. It would follow that we need only to ban stampedes to address this issue. But this line of reasoning neglects an important aspect of consideration—whether there is time to debate and discuss the utterance. A few seconds can be the difference between life and death

when escaping from a burning building, and as such the situation does not afford adequate time to verify that there really is a fire. It is entirely reasonable to expect that people will panic and cause a stampede when suddenly told of an imminent threat to their lives. Moreover, there is no perceptible social benefit or justification for shouting fire in a crowded theater beyond wanton exercise of speech rights, and for these reasons it makes more sense to hold the speaker accountable for shouting than it does to hold the crowd accountable for its stampede.

Whether you agree or disagree with the court's decision in *Schenck v. United States*, it should make clear that while the harm principle informs where we place our limits on free speech, it is up to the courts at any given time to decide what constitutes a clear and present danger. If the court decides that the danger is clear and present enough, speech can be suppressed, even if it is out of concern for a seemingly abstract threat. Conversely, speech which does pose a threat is nevertheless protected as long as it remains vague or not pressing (such as a call for revolution in the indefinite future).

This means that the judges' beliefs factor greatly into the outcome of a case, since those beliefs are what the judges use to determine what constitutes a "clear and present" danger. Justice Wendell Holmes stuck to the interpretation he established in *Schenck v. United States*, and his choice to dissent in a similar case helped to clarify that standard. In *Abrams v. United States,* 250 U. S. 616 (1919), the defendants distributed leaflets denouncing the sending of United States troops to Russia. The majority ruled against the defendants, but Justices Holmes and Brandeis dissented, arguing that the defendants in *Abrams* had not attempted to actively disrupt the war effort like Schenck and Bauer, that the pamphlets therefore fell short of the clear and present danger standard established by *Schenck v. United States*, and that the defendants

were really being persecuted for their ideas, not because they were a real danger to the nation or its war effort. In Holmes' own words,

> "It is said that this manifesto was more than a theory, that it was an incitement. Every idea is an incitement. It offers itself for belief and if believed it is acted on unless some other belief outweighs it or some failure of energy stifles the movement at its birth. The only difference between an expression of opinion and an incitement in the narrower sense is the speaker's enthusiasm for the result. Eloquence may set fire to reason, but, whatever may be thought of the redundant discourse before us, it had no chance of starting a present conflagration. If, in the long run, the beliefs expressed in proletarian dictatorship are destined to be accepted by the dominant forces of the community, the only meaning of free speech is that they should be given their chance and have their way."

The Supreme Court would continue to rule on speech cases using the "bad tendency test" until *Brandenburg v. Ohio,* 395 U. S. 444 (1969). This occurred most notably in *Whitney* v. *California,* 274 U. S. 357 (1927), a case which really helped to solidify it. Charlotte Anita Whitney was a founding member of the Communist Labor Party of California, and in 1927 she was prosecuted under the California Criminal Syndicalism Act of 1919. While Whitney made some other additional arguments, the case really hinged on her claim that California's Criminal Syndicalism Act violated the First and/or Fourteenth Amendments. The Court ruled against her, and held that the government may use its police power to punish those who abuse their freedom of speech "by utterances inimical to the public welfare, tending to incite crime, disturb the public peace, or endanger the foundations of organized government and threaten its overthrow." This effectively defined the bad tendency test. On the surface it seems reasonable enough, but if we remember the previous example of Schenck and factor in consideration of the Red Scare which was ongoing at the time, it is easy to see how problematic the bad tendency test really is. The test allows the state to ban or punish speech of any kind, provided it can be construed as meeting the aforementioned criteria. As one of those criteria is speech which "endanger[s] the foundations of organized government and threaten[s] its overthrow," any

ideology or set of ideas can be justifiably suppressed as long as it is deemed sufficiently dangerous. Again, while the criteria seem fairly reasonable and all appear to cover problematic exercises of speech which nobody should reasonably demand the right to engage in, don't forget that this was essentially the same court that deemed the printing of anti-war pamphlets a serious threat to national security. In other words, we should not assume that other people in different socio-political contexts and with different biases will interpret the criteria in the same way that we do. The bad tendency test thus fails as a free-speech principle because its criteria are too vague and thus give the courts near total power to decide which speech is a threat. Clearly, a better definition would be needed.

This is where the significance of the *Brandenburg v. Ohio* decision becomes apparent. Clarence Brandenburg was a Ku Klux Klan leader in rural Ohio, who invited a reporter from a Cincinnati news station to cover his Klan rally. During the rally, Brandenburg advocated for violence against blacks, jews, and those who support them, and announced plans for a march on the Capitol on the fourth of July. Brandenburg was charged with advocating violence under Ohio's Criminal Syndicalism statute, which, like the other criminal syndicalism acts, forbade even abstract calls to violence. As the bad tendency test made no distinction between likely and unlikely harm, Brandenburg quickly lost his initial court case, but on appeal the Supreme Court overturned the previously established bad tendency test, finding that abstract threats or calls for violence were still protected by the First Amendment. While Brandenburg had advocated for violent action, his targets were large, abstract groups (Blacks, Jews, and their allies) and his calls for violence did not include a target date or time, making it hard to argue that he or his followers posed a credible, imminent threat of harm or lawless action. Unlike earlier definitions of the limits of speech, this new standard drew a clear distinction between the abstract and concrete

dangers posed by a given utterance. Applying this standard to *Schenck*, we might agree with the court's assessment that the ideological content of Schenck's pamphlets posed a threat to the state in the abstract, but those pamphlets could hardly be considered a concrete imminent danger to state security, and would therefore fall short of being a threat under the new standard. Likewise, even if I claim to be a radical anarchist and insist that I do not believe in the values enshrined in the Constitution, that does not mean I pose any real danger to the state—I am just one person with limited resources and no following. In other words, the test was now two-fold: speech must be "directed to inciting or producing imminent lawless action" *and* "likely to incite or produce such action."

It is worth mentioning *Near v. Minnesota*, 283 US 697 (1931), in which the Supreme Court struck down a proposed Minnesota law which would have prescriptively limited press publications by repeat offenders in a manner similar to the bad tendency test. Jay Near was the editor of *The Saturday Press*, and a frequent muckraker who earned the ire of Minnesota state officials when he published a series of defamatory articles questioning the quality of their public service in 1927. The state of Minnesota responded by filing an injunction barring publication of *The Saturday Press* under Minnesota's Public Nuisance Law of 1925 (also known as the "Minnesota Gag Law"). This law forbade *The Saturday Press* or any other publication which had been deemed a public nuisance from publishing. The Supreme Court found the provision facially unconstitutional, deciding that while the government is allowed to punish speech after the fact under certain conditions, it is never allowed to prescriptively limit what can be said or published. So while Minnesota might have had the right to punish press organizations for harm they caused after the fact, it can never prevent them from publishing anything in particular.

As press organizations have more following and influence than a given individual, you would think that they would be more likely to be held liable for damage caused by their speech. This has occurred in some cases (a recent example being *Bollea v. Gawker*). But a series of significant freedom of the press cases has established as precedent that the press serves an important function in the role of government accountability, and therefore enjoys strong protections regarding what it is allowed to publish. This perspective often accompanies the view that the press acts as a sort of "fourth institution" of government, creating an avenue by which whistleblowers can reliably publish important information that has been kept from the public.

A good example of this phenomenon is the case of the Pentagon Papers. In 1971, Daniel Ellsworth leaked a classified report assessing the goals, costs, and objectives of the Vietnam war to the *New York Times*. Ellsworth's mitigating circumstances as a whistleblower are worth noting here: while the report was classified, it mostly focused on previous events in the war (specifically up to 1968), and Ellsworth took great care to exclude details he deemed relevant to national security (such as the names of current operatives or ongoing projects). The Nixon administration attempted to block publication of the report on those grounds, but the Supreme Court ultimately ruled that the First Amendment necessitates a "heavy presumption against" prior restraint, and that the government was unable to provide substantial proof that publication of the papers would place American forces in "inevitable, direct, and immediate" danger. In other words, while national security is a legitimate reason for suppressing speech, the danger posed by a given utterance must be substantiated in a similar manner to the test presented in *Brandenburg v. Ohio*.

Aside from issues of national security, the press sometimes also faces pressure from libel suits, which are meant to redress harm caused by slander. To give an example, by publishing an article claiming that you are a pedophile, even if there is no truth whatsoever to it, I could

damage your public reputation massively. This is all the more true with the rise of the internet, since a simple Google search of one's name is enough to bring up any slander about them (I will discuss this in more detail in chapter 3). Libel laws allow people who feel they have been greatly wronged by slander to sue for remuneration, with some caveats. *Hustler Magazine v. Falwell*, 485 U. S. 46 (1988) brought this issue under the judicial microscope, when fundamentalist minister Jerry Falwell sued *Hustler Magazine* for publishing a highly offensive parody of an ad campaign, which proclaimed that Falwell has drunken incestuous sex with his mother in an outhouse. Falwell won the initial case, but on appeal the Supreme Court found that it could not produce an adequate definition of parody which would not also have chilling effects on freedom of expression, and that the state's interest in protecting free expression surpasses its interest in protecting public figures from offense. Hence, barring extraordinary circumstances, the act of parody is protected under the First Amendment.

It is also worth mentioning *New York Times Co. v. Sullivan*, 376 US 254 (1964) as it established an additional standard of scrutiny for public figures in libel cases. Prior to this case, public officials in the South had been using libel suits to obstruct media coverage of civil rights protests. *New York Times Co. v. Sullivan* found that public officials suing for libel must demonstrate not only that the information published about them was false and damaging, but that it was published with deliberate malicious intent. In other words, even factual inaccuracies must be demonstrated to be the product either of malicious intent or reckless negligence for a public official to prove libel in court. This decision is even more significant in light of the financial burden lawsuits impose. A harsh damages settlement can bankrupt a smaller press organization (e.g. *Curtis Publishing Co. v. Butts* 388 US 130 (1967), *Bollea v. Gawker* 170 So.3d 125 (2015)) and this can create an incentive for press organizations to self-censor and avoid offense out of

fear of legal action being taken against them. *New York Times Co. v. Sullivan* thus strongly cements freedom of the press and limits the applicability of libel laws to private citizens (as opposed to media organizations) in most cases.

We now turn to cases which dealt with the First Amendment's content-neutrality. Under our current interpretation of the First Amendment, no law may abridge freedom of expression on the basis of the content of that expression or disagreement with the ideas expressed. Furthermore, any restrictions on expression must pass strict scrutiny and be narrowly tailored to serve a compelling state interest. A law can be deemed content-based even if it is facially neutral. In the example of *Ward v. Rock Against Racism* 491 US 781 (1989), the Supreme Court found that as long as the "means chosen are not substantially broader than necessary to achieve the government's interest," a regulation cannot be found invalid simply because there is a less intrusive way to address the issue. In other words, even if there is a less intrusive or non-regulatory solution, a regulation can still be allowed to stand if it achieves the government interest much more effectively than alternative solutions. Likewise, policies like affirmative action are upheld because they are narrowly tailored to serve a compelling state interest (increasing student-body diversity on university campuses) and accomplish that interest much more effectively than non-regulatory means. If either of these factors ceases to hold true, affirmative action would either need to be justified on other grounds, or would cease to be constitutional.

Laws which are facially content-based are always subject to the strict scrutiny test, but may still be allowed if they are narrowly tailored to serve a compelling state interest. It is rare, however, that a law meets this bar: many local laws aimed at addressing a perceived public nuisance have been struck down on the basis that they are content-based and either insufficiently

compelling or overly broad. These include such laws as a Gilbert, Arizona ordinance regulating the placement and display of public signs, which was struck down for including categories which distinguished between temporary directional, political, and ideological signs.[14] Many state-level attempts at regulating hate speech have failed for similar reasons, including the notable example of *R.A.V v. City of St. Paul,* 505 US 377 (1992). This case primarily dealt with a Minnesota statute barring expressions of hatred on the basis of race, ethnicity, gender, etc. and whether that statute violated the First Amendment. The defendant was a minor at the time and is therefore referred to as "R.A.V." R.A.V. and his friends built a crude wooden cross out of broken chair legs and burned it on the lawn of R.A.V.'s black neighbors across the street. The law was initially deemed constitutional by equating its limitations with the notion of "fighting words" established previously in *Chaplinsky v. New Hampshire*, 315 US 568 (1942). But while the court condemned R.A.V.'s actions, it ultimately nevertheless struck down the Minnesota law, because it prohibited discrimination on the basis of *certain* traits and/or group membership, rather than others, and this in the Court's view constituted a content distinction, making the law content-based and thus facially unconstitutional. In the words of the Court, government has no authority to "to license one side of a debate to fight freestyle, while requiring the other to follow the Marquis of Queensbury Rules."[15] While the law could have still been justified by means of a compelling state interest, the Court also found that Minnesota already had enough laws in place to prevent behavior such as R.A.V.'s without appealing to the hate-speech prohibition, negating any need-based justification for the statute's continued enforcement.[16]

[14] *Reed* v. *Town of Gilbert,* 576 U. S. 155 (2015)
[15] "R.A.V. v. City of St. Paul." Oyez. Accessed May 8, 2022. https://www.oyez.org/cases/1991/90-7675.
[16] Ibid

The Court tends to be similarly unsympathetic towards attempts to ban speech on the basis of offense, however offensive the speech in question may be. This is largely out of a sense of guarded caution against the erosion of First Amendment principles, but also stems from the acknowledgement that many kinds of speech can be deemed offensive depending on the individual, and that the bar for speech to be considered truly offensive is accordingly set very high. Municipalities need the power to address public nuisances, but such restrictions can seriously limit free expression, or even be an instrument through which the majority can silence dissent. Generally speaking, some form of Feinberg's Offense Principle is applied. Offense which an individual can easily avoid (such as a book, which one can simply close and set down) constitutes less of a public nuisance than pornographic billboards on the highway. While the Court has generally adopted a "know-it-when-you-see-it" stance on obscenity, the exact limits and technical definition of the obscene remain subject to debate. The closest thing we have to a definition is something called the Miller test, which was established by the precedent of *Miller v. California,* 413 US 15 (1973). The Miller test stipulates that content is offensive if it meets all three of the following criteria: (a) whether "the average person, applying contemporary community standards" would find that the work, taken as a whole, appeals to the prurient interest, (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value."

There is evidently a need to allow municipal governments to address public nuisances and other problems caused by wanton abuse of speech, but at the same time our inability to produce a satisfactory definition of dangerous or sufficiently nuisance-causing speech makes it nigh-impossible to craft a law capable of surviving strict scrutiny. *Virginia v. Black*,

538 US 343 (2003) is a good example, since it deals with a Virginia statute prohibiting the burning of crosses, a well-established symbol of hate. Unlike the Minnesota statute under scrutiny in the R.A.V. case, the Virginia law narrowly targeted the act of burning crosses. Although the Supreme Court found that the law was at face constitutional (equating cross-burning with intent to intimidate with the legal notion of a "true threat"), since the Virginia law assumed all cross-burning is done with the intent to intimidate, this had the consequence of placing the burden of proof on the defendant that the burning was not done with intent to intimidate, and the Supreme Court found this to be in violation of the Fourteenth Amendment and therefore unconstitutional (as it unduly influenced how the jury was to look at the case). Justice Clarence Thomas dissented arguing that cross-burning should itself be a category of excepted speech, given its long history and association with terrorism, but his was the minority opinion.

United States court precedent makes clear that any sort of established political "truth" (or orthodoxy) is principally disallowed by the First Amendment. This is perhaps best crystallized in Judge Timothy Black's decision on *Susan B. Anthony List v. Ohio Elections Commission*, 134 S.Ct. 2334 (2014):

> "Lies have no place in the political arena and serve no purpose other than to undermine the integrity of the democratic process. The problem is that, at times, there is no clear way to determine whether a political statement is a lie or the truth. What is certain, however, is that *we do not want the Government deciding what is political truth* — for fear that the Government might persecute those who criticize it. Instead, in a democracy, *the voters should decide.*"

I imagine few people would disagree with this statement. Yet there remains a problem. As history (especially *recent history*) demonstrates, just about any issue can become political—even matters on which there is overwhelming scientific consensus, such as climate change or vaccination. We

tend to think that we can separate these fields; that it is possible to have scientific consensus whilst respecting our differences of political opinion. But in a world where such boundaries have become inextricably blurred—one we increasingly find ourselves living in today—the political is unbounded. Any issue has the potential to become politicized, in turn precluding any consensus on its basic facts. In other words, the First Amendment's prohibition on any kind of political orthodoxy *necessarily also* prohibits *any* kind of orthodoxy whatsoever, as no clear distinction can be made between the political and apolitical. Given the inherent limitations of human knowledge and science's tendency to overturn periodically its underlying assumptions, we cannot treat it as the be-all end-all of objective truth, nor can we ignore the assertion that scientific knowledge is just another example of human consensus, albeit one strongly grounded in empirical evidence. We can however respond that scientific knowledge and the consensus it offers on basic facts about our physical reality are incredibly useful. We care very much, for instance, that honey is toxic for infants and young children, that lead and arsenic are poisonous when accumulated in the body, that antibiotics are effective at treating bacterial infections, etc. because being wrong about these facts has potentially fatal consequences. Likewise, we *should* have good reason to care that vaccines are effective at treating viral infections, or that carbon emissions result in climate change in the aggregate. There is strong scientific consensus on all of these facts, yet the latter two are subject to enormous political controversy. As the existence of the Flat Earth Society makes abundantly clear, no fact about our reality, however apparent or self-evident it may appear to be, is beyond politicization.

Yet this peculiar inconsistency with how we treat scientific information persists. On the one hand, we see no problem with scientific orthodoxy when a judge calls on expert witnesses in court, or when the President appoints scientific consultants. Yet we are deeply squeamish about

asserting scientific consensus on anything we feel has been touched by political debate, as suddenly the issue is felt to be too big for any institution to decide and must be left to the discretion of the general (voting) public. To quote Susan B. Anthony List in the aforementioned 2014 case, "[We are not] arguing for the right to lie. We're arguing that we have a right not to have the truth of our political statements be judged by the government." The problem is that politicization can mask and brush over basic factual inaccuracies, even in cases where it actually *is* possible to confirm the truth or falsity of a given statement. While we certainly don't want the government to decide (let alone mandate) political truth, it is also the case that we want some agreement on basic non-political truth, and in the event that all truth becomes political truth, we are principally forbidden from ever officializing such an agreement, and must perpetually do battle against disinformation in the realm of public discourse, all the while bearing the costs that disinformation imposes upon the public.

I do not share Mill's optimism that the free marketplace of ideas will allow us to muddle through, nor do I think we should resign ourselves to suffer the costs of disinformation when there are measures we can take to mitigate its damages. It is tempting to dismiss the notion of costs as avoidable because autonomy always includes the right to harm oneself, but we should not be so quick to do so when those costs come in the form of human lives (as COVID has demonstrated). Given how uninformed the average voter is on most policy issues, coupled with the strongly uneven distribution of speaking power within public discourse in the United States, it is hard to imagine that popular opinion will pay much heed to expert consensus when it is more expedient for people to ignore it. Moreover, in a world where every truth is politicized, it becomes possible for candidates to market their own versions of reality (like QAnon) to voters, who in a postmodern incarnation of Mill's marketplace of ideas pick out their favorite and most comfortable truths from

the various salesmen hawking them. As markets only are concerned about truth when its absence reduces profit, it is hard to imagine salesmen in such a marketplace would pay much heed to the truthfulness or accuracy of the "facts" they are peddling. In such a marketplace, truth is a function of expedience, and what is expedient is not necessarily what is best.

Chapter 3—Free Speech in the Digital Age

The internet currently has no laws governing expression. There are no protections for speech, nor are there prohibitions on what may be said online (with the exception of copyrighted material and child pornography, which are handled by the courts and police/FBI, respectively). Instead, everything is left to the discretion of private websites, which moderate speech as they please. So far, this has generally worked out: there are enough websites that the internet can behave like Mill's free marketplace of ideas. It isn't a problem if websites ban particular kinds of speech, because those who disagree with the prohibitions can simply voice their opinions elsewhere. In theory, the prevalence of certain ideas should therefore naturally indicate their merit: moderation strategies which work well will be more commonplace, and speech which is considered a major nuisance which be pushed to the fringes, and appear less often. There isn't a need to protect strongly such speech, because the internet is considered a private domain and de-platformed speakers can typically find another platform willing to host them, or even make their own. Combined with all the difficulties that arise when you try to produce an adequate definition of unacceptable speech, the web makes for a pretty compelling argument to leave free speech as it is.

This is especially the case when contrasted with China's approach to internet censorship as embodied in the Great Firewall, wherein censorship enforcement is outsourced to the private

companies hosting internet platforms, and the government prescribes what can and cannot be said. Some have argued that the decentralization of the internet renders most attempts at censorship moot, but even China alone is enough to demonstrate that this claim is unfounded. Because internet access is most often provided by private businesses (Internet Service Providers, or ISPs), it is possible to control a domestic population's internet access by imposing restrictions on what ISPs are allowed to provide access to or otherwise coercing them into cooperation. In areas where municipal broadband is used, the government can skip this step as it directly controls internet access. Even assuming a state of net neutrality and open internet access, content hosting sites can still be restricted resulting in a similar degree of censorship. This is precisely the reason that content hosts were given freedom from liability for user-generated content in section 230 of the Communications Decency Act (CDA) of 1996: there was widespread concern that such liability would make hosting user content too risky and thus result in a chilling effect on online speech. Reversing this logic, we can see that China has utilized this effect to privatize its censorship enforcement by making content-hosting companies liable for user breaches of censorship policy. This provides a powerful incentive for companies to self-censor as no user speech is worth making an enemy of the government. In addition, the ability to circumvent digital censorship measures itself requires some amount of technological literacy. If you want to learn how to use a VPN, a simple Google search will suffice, providing you with hundreds of detailed tutorials and ranked lists comparing VPN services by features and pricing. In China, where the very word "VPN" is banned and there is no access to Google (Baidu cannot display banned results), this constitutes a sort of Catch-22, wherein it is difficult for most people to even acquire the prerequisite knowledge to begin circumventing censorship. It is thus a mistake to think that digital censorship regimes of this sort are too costly and impractical to maintain.

Which is not to say that we should feel totally comfortable with how we treat the internet in the United States—because circumstances have changed. The internet can no longer be considered a separate space divorced from real world consequences and populated mostly by subcultures and technology buffs isolated from mainstream society. It now has come to dominate our everyday existence. Popular public forums have moved from the town square or local pub to social media sites like Meta (formerly Facebook), Twitter, Instagram, YouTube, Reddit, NextDoor, etc. Most print news media now also (or even primarily) publish in digital formats, many of which have comment sections where people discuss current events. Nearly every elected U.S. politician has a Twitter account, and most use other social media platforms as well. And this is without mentioning the countless discussion boards and forums out there, many of which are devoted to particular hobbies or niche interests. In short: most of our public discourse has moved online, and it is impossible to ignore the internet and the influence it has on society.

How we look at speech online has not kept up with these developments. As mentioned previously, the free marketplace of ideas requires an assumption of faith in the wisdom of the crowd, and rests on certain premises regarding the form and nature of discourse as well as how conclusions are reached, such as the assumption that ideas will eventually lose public support once refuted. Like any market mechanism, the internet rewards popularity and demand, not a particular set of normative values. And like any market mechanism, if we wish to ensure that it produces good outcomes, we must occasionally steer it towards them. It is easy to compare it to a mechanism like consensus, but this would be a conflation—consensus factors in expert opinion and evidence, and is no mere popularity contest. Rather, it is because we still care about what is true and what is right even in the face of overwhelming public opposition that we must draw this distinction. If a majoritarian popular opinion alone was enough to assert an idea as objective

truth, our constitution would not have such strong protections for minority rights or the right to dissent. In other words, it is a mistake to think that the popular vetting of an idea is tantamount to a social, much less political, consensus on its worth or truthfulness.

My primary concern is with disinformation, because the internet's circumstances uniquely undermine and bypass normal barriers and checks against disinformation. Recall previous concerns about the printing press. While it was true that the invention of moveable type in the 15th century made it much easier to publish and disseminate ideas, it still required a large investment of time and resources to operate and maintain. In other words, only print shops with the funds and the expertise to own and operate printing equipment would be able fully to utilize the new press, which is why centers of wealth like Venice became dominant in the trade. Anyone who desired to open up his own printing shop would need a considerable amount of capital to invest. So while moveable type democratized publishing power to an extent, there were still significant barriers to entry. These constituted *de facto* gatekeepers and material checks to publishing, although they were largely insufficient in sectarian Europe. The Roman Catholic Church developed the *Index Expurgatorius* to check publication of works deemed heretical or even dangerous. In jurisdictions where the church had influence, books could not be printed without ecclesiastical approval. The infamous English Licensing of the Press Act of 1662 is another good example, as it similarly attempted to limit the publication of materials deemed dangerous to the state by restricting who could own and operate a print shop. With a list of registered print shops, it would theoretically be possible to locate even the printer of an anonymous pamphlet (and eventually find the author), since the list should account for all print shops capable of printing the pamphlet. Milton saw prepublication censorship as unnecessary for precisely this reason: "…as for regulating the Presse, let no man think to have the honour of

advising ye better then [Parliament] have done in that Order publisht next before this, that no book be Printed, unlesse the Printers and the Authors name, or at least the Printers be register'd. Those which otherwise come forth, if they be found mischievous and libellous, the fire and the executioner will be the timeliest and the most effectuall remedy, that mans prevention can use."[17]

The internet does away with all such barriers. It provides in one package the stage, audience, and tools needed to deliver a message to them. Anyone can host a website for free and with relatively little tech savvy, and there are few regulatory constraints. In recent years this has blurred any distinction between traditional established news sources with reputations and in most cases a commitment to journalistic rigor and newer digital publications with less commitment to objective reporting.

The internet also introduces another variable into the mix: anonymity. It is not as if anonymous publishing is anything new, but authors have often historically been persecuted for their more radical ideas. Hence the use of aliases, pseudonyms, and other devices to hide identity. What the internet offers that is new is the degree to which anonymity can be preserved. Nobody knows who Satoshi Nakamoto, the mysterious creator of bitcoin, is. While many have come forward claiming to be him, only a message using the original/real Nakamoto's PGP key can prove his identity (via a special cryptography technique), and so far, none have appeared.[18] While it is possible for government and other agencies to crack such encryptions (at least in the case of standard security measures like Tor or TAILS Linux), it requires a considerable amount of time and effort, especially when compared to the ease with which such encryption methods can be employed. Anonymity also nullifies the social costs of rumormongering or spreading

[17] Milton, *Areopagitica.*
[18] Kay, "The Many Alleged Identities of Bitcoin's Mysterious Creator, Satoshi Nakamoto."

disinformation, as there is no reputation attached to anonymous utterances and thus no consequences or accountability incentives for the speaker. With older methods of communication, individuals who repeatedly spread false information over time developed reputations for being untrustworthy or unreliable and faced reprobation for their behavior, like the boy who cried wolf. People were discouraged from making baseless claims and encouraged to pursue factual accuracy, or at least were motivated to provide *useful* information. In short, the internet is very different. It makes possible the transmission of ideas on a never-before-seen scale anonymously, such that millions can be reached with relatively few barriers, and every idea will find *some* audience. Factor in the aforementioned blurring of news sources and the widespread absence of journalistic rigor on many of them, and suddenly almost any idea becomes salient. In this new, virtual space, nearly any idea can take hold, whether it is properly researched, argued intelligently, and conducted responsibly or not.

I am certainly not alone in thinking this. In his 2018 book *Minds Make Societies*, Pascal Boyer examines the function of so-called "junk culture"—information which serves no apparent purpose yet often still manages to spread widely—and concludes that the internet greatly facilitates its spread and transmission. He sees two main reasons for this:

> "First, as we all know, connectivity makes it cheap to acquire information and makes the cost of broadcasting almost negligible. That is not just because connections are cheap but also because the role of reputation is greatly diminished. Again, consider a small-scale society. In such a group, accusations of witchcraft, for instance, are potentially very costly. You never know for sure that people will not rally around the alleged witch. You may pay dearly if you are the only one to level the charge against a particular individual. That is why public accusations of this kind only occur after a long period of discreet consultations, and in some places are never made public. **By contrast, modern connectivity allows both anonymity and geographical distance, virtually eliminating the social costs of accusations and rumormongering**. So it is no surprise that Internet rumors and crusades are so vicious in tone, so quick to emerge, and often so expansive.

> Second, worldwide connectivity may fuel our worst dispositions to create and broadcast junk culture, by fooling us into illusions of consensus. Consider this. In a small-scale society or in a village, those who come up with some new variety of, say, magical beliefs, will probably find very few people who share their strange notions. By contrast, in a connected world that includes billions of users, almost any kind of harebrained proposition is probably already promoted by thousands of individuals or more. **So connectivity is likely to provide all with an inflated sense of consensus around their own ideas**—a propensity that was already observed in experimental studies. This effect could be multiplied by the illusion that sources are independent. That is to say, if we find out that thousands agree with us, for example, that the world is indeed controlled by alien reptiles, we marvel at the fact that so many great minds think alike. It would seem that all these individuals independently converged on that same theory, when in all likelihood many of them read the exact same web page."[19]

These observations undermine the assumptions behind Mill's free marketplace of ideas, insofar as the marketplace model presumes that there are reputational costs to making false allegations and that consensus on an issue can only form once there is a majority opinion on it within the broader community of public discourse. The former observation is largely self-evident from online discourse, but perhaps its most potent manifestation is *trolling*, which is entirely made possible by internet anonymity. Trolling is essentially the act (or art, depending on who you ask) of tricking people into getting mad and wasting their time arguing with you. Trolls make deliberately inflammatory remarks to provoke a particular (usually angry) response from their target. Often this involves convincing the other person that you genuinely hold a highly dubious or otherwise controversial view, which they will feel obligated to respond to and/or debunk. As the troll's only real objective is to provoke anger (most often for his own amusement), his target has no hope of actually convincing the troll to change his position and is essentially wasting his time arguing.

Trolling is made possible by the internet and its anonymity because it is otherwise difficult to engage in such behavior without facing substantial social reproachment. Trolling by

[19] Boyer, *Minds Make Societies*, 88–89.

itself may not seem like much of a concern, but we must remember that it is an instantiation of the broader phenomenon of disinformation. During the 2016 Presidential election, trolling was briefly popularized within public discourse by the revelation that the Russian state had employed professional trolls to spread election-related disinformation. We might not be overly concerned about people looking for a laugh at others' expense, but we have much more reason to be concerned with foreign state actors seeking to exploit the vulnerabilities of open societies. As there is no way to infer reliably a speaker's motives or whether their belief is genuine, it becomes relatively easy for agents of discursive power to infiltrate the domestic crowd and add their voices to the discourse.

The latter of Boyer's observations is by no means new, and is known to psychologists as the problem of false consensus, but the internet greatly amplifies this tendency by providing access to audiences unprecedented in scope and scale. False consensus can allow even the most thoroughly debunked assertions to survive public scrutiny provided there is a sufficient number of devout adherents. Discredited British physician Andrew Wakefield's claims that vaccines caused autism in children is a good example of this phenomenon, as despite universal debunking the belief continues to linger in the public sphere. It is also worth considering the example of so-called "zombie statistics"—statistics which have already been debunked/disproven yet continue to be quoted and cited popularly. Notable examples include "People only use 10% of their brains" (as seen in films like *Lucy* and *Limitless*), "Drinking 8 glasses of water a day is good for your health," and "Women provide 66% of the work, produce 50% of the food, but earn only 10% of the income and own 1% of the property."[20] These zombie statistics are given further life by an abundance of contested or misleading statistics, which have some truth to them but present

[20] Kodan, "🧟 Zombie Statistics."

misleading conclusions that do not follow strictly from evidence. For instance, the statistic claiming that "the 20th century was the bloodiest in human history" is true if "bloodiest" is defined as highest number of casualties, but ceases to be true when the relative size of the population is taken into consideration (the title of bloodiest then actually belongs to the 8th century). Similarly, the statistic that women earn 78 cents on the dollar is true when considering just the salaries earned by men and women respectively but does not account for differences in career choice and/or hours worked, and once these are controlled for the gap narrows to 95 cents on the dollar.[21] The prevalence of such statistics makes it easy to assume that other statistics, like the one about women owning 1% of the world's property, must similarly have some element of truth to them (in fact, this particular statistic stems from a 1978 publication by the International Labor Organization which provided no evidence for these claims), and adds to our predisposition to accept propositions as true on the basis of plausibility alone. These statistics can make for a powerful call to action, but they can also obfuscate the terms of the debate, and rarely does the desire for factual precision outweigh these statistics' usefulness in recruiting new believers to a cause.[22]

This is where the problem of disinformation starts to become so apparent. As a democracy, we are obligated to care about the nature and quality of our factual discourse. We rely on the "wisdom of the crowd" to ultimately make the correct choices, and to factor in expert opinions without being overwhelmed by them. This is a delicate and difficult balancing act for anyone to perform, and the internet offers unadulterated access to an entire sea of information, with no guidance on how to filter or view it. As a result, the burden of sorting fact from fiction

[21] Chamberlain, Zhao, and Stansell, "Progress on the Gender Pay Gap: 2019."
[22] Mangu-Ward, "Zombie Statistics."

and evaluating sources to find the truth falls squarely on the reader/viewer. Content-recommending AI can perpetuate this problem by ushering in fallacious material and creating user-tailored echo chambers. We can combat this by modifying and regulating how these algorithms work (for instance, Facebook recently opted to remove political news from content-suggestion).[23] But a real challenge remains: how to tackle the organic spread of disinformation that has dire consequences.

This is incredibly difficult, if not nigh-impossible to do, without abridging freedom. Let us consider two attempts at preventing children from accessing explicit web content in the late 1990's. Both resulted in the Court striking down proposed legislation for being overly broad and/or overly restrictive. In *Reno v. American Civil Liberties Union* (1997), the Court struck down provisions of the Communications Decency Act (1996) which would have criminalized the intentional transmission of "obscene or indecent" messages to minors, on the basis that the law failed to adequately define "indecent" communications. In *Mainstream Loudoun, et al. v. Board of Trustees of the Loudoun County Library* (1998) the Court struck down a law requiring public libraries in Loudoun County, Virginia, to have web filtering software to prevent access to pornography on the basis that it was unduly restrictive. But only a few years later, in *United States v. American Library Association* (2003) the Court upheld a similar law requiring that all public libraries install filtering software to continue receiving federal funding. The main difference between these cases is the ease with which restricted sites can be unblocked. In *Loudoun*, petitioners needed to file a written request with the library, including the reason for unblocking a site, and the library would only unblock it after review. But the policy in the 2003

[23] Roose and Isaac, "Facebook Dials Down the Politics for Users."

case allowed librarians to unblock a site immediately for any adult on request, making it comparatively less restrictive.

These cases are useful when considering how to create a disinformation law which doesn't unduly restrict freedom. As evidenced by *Reno*, any definition of disinformation used in such a law would need to be atomically precise. And as demonstrated by *Loudoun* and *American Library Association*, any prior restraint would need to be easily overcome by legitimate speakers, as otherwise it would constitute an undue burden on everyone's freedom of speech. Since there is no easy way to differentiate between legitimate and illegitimate speech (unlike the ease with which we can differentiate between children and adults) it is highly unlikely that any prior restraint on speech whatsoever would pass muster.

Setting aside prior restraint, it is worth considering what constitutes a compelling state interest. In the cases of *Loudoun* and *ALA*, the government agreed that protecting minors and preventing the public display of elicit and highly offensive material are both compelling state interests. But what about disinformation? Can we justify prohibitions on speech merely by reference to a potential bad outcome? Can we argue that the eventual damage caused by unchecked disinformation justifies its early prohibition? Disinformation is too broad a category to argue that the act itself is fundamentally harmful, since the spreading disinformation encompasses a wide range of activities. Disinformation could be a salesman hawking fake COVID-19 cures, propaganda by a foreign state actor, or even a complete "alternate" reality like QAnon, with its own set of tenets and truths.[24] We can help to categorize and differentiate cases based on the severity of the harm, audience outreach, likelihood of follow-through, etc. but as we can observe from previous cases, abstract danger is not enough; conversely, we already have

[24] Warzel, "Is QAnon the Most Dangerous Conspiracy Theory of the 21st Century?"

laws concerned with concrete threats posed by speech. Someone who encourages children to drink bleach can likely be held liable if those children obey and die. But abstract threats, however looming, forbid immediate action. The kind of disinformation we're worried about is the kind powerful enough to undermine the institutions of government. But for any information to gain such a level of widespread acceptance means that a lot of people must consider it truth. And as we have discussed, the government does not and arguably should not consider itself to be in the business of legislating truth.

Clearly then, if there is a governmental solution to this problem, it would have to take the form of a committee, agency, or special designation for websites, as no law can be constructed which would not majorly hinder everyone's freedom. A committee could overcome the limitations of a law by being context aware, but any such institution only raises more questions of accountability, and an overzealous committee could easily come to see its job as prescribing the truth. Since we are a society in which the truth is determined by popular majority ("let the voters decide"), this may mean some ugly yet unavoidable outcomes wherein we endure costs to human life and welfare while sticking to whatever truth the voters decide, even when their decision is based on false premises.

There has been much discussion of "cancel culture" in the last decade or so. For those unfamiliar with the term, cancel culture refers to the social costs one must pay when going against the orthodoxy of political correctness. Conservatives have often complained that they cannot express their views without facing considerable backlash from defenders of this orthodoxy, and this can result in loss of one's livelihood and reputation. A number of senior university professors have faced cancellation, some losing their positions. Of course, freedom of speech does not mean freedom from consequences, but it is nevertheless true that we are

obligated to at least consider the chilling effects this may have on speech. Some have even argued that social media needs to be subject to the same First Amendment protections for speech that apply offline. It is arguably highly impractical to impose such requirements on websites universally, however, as sites rely on a great deal of latitude in dealing with speech in order to make content moderation practicable. For instance, suppose we were to require that all speech be allowed on all websites barring violation of the harm principle. How would websites deal with spam? We could attempt to make an exemption for spam, but to do so we would first require a precise and satisfactory definition of what spam is. You could argue that if "know it when you see it" is enough for our legal definition of obscenity, a similar formulation should be sufficient for spam. At the same time, it is hard to miss the potential such a vague definition has for abuse. Like local communities, sites will always have a need to engage in some moderation of speech, and thus the real questions are who should decide, how such decisions are to be made, and how their makers are to be held accountable.

Given the practical difficulties of attempting to enforce First Amendment rights online, it is implausible to impose a blanket requirement. I do, however, see considerable merit in a sort of too-big-to-fail designation for large platforms important to public discourse, like Facebook, Twitter, and Reddit. This would render them subject to additional requirements in their treatment of speech which other smaller sites would not have to follow. Such a distinction would allow for flexibility on the part of smaller sites and communities, whilst attempting to preserve speech on larger platforms. Such requirements could demand deference to the First Amendment when assessing speech instances, preventing censorship on ideological grounds, and would simultaneously indicate the officiality or importance of large public platforms, by drawing a line between those platforms which are subject to the requirements and those that are not.

A better option would be to roll back section 230 of the CDA. This may seem like an extreme measure at face value, but the Department of Justice has already acknowledged problems with the current interpretation and treatment of speech online. In a 2020 review it conducted of section 230, it found that "[t]he combination of significant technological changes since 1996 and the expansive interpretation that courts have given Section 230, however, has left online platforms both immune for a wide array of illicit activity on their services and free to moderate content with little transparency or accountability."[25] While it is true that section 230 was written out of a concern for the chilling effect liability could have on speech online, this happened at a time when the internet was still in its early stages and had yet to play a large role in public discourse. Given the explosive growth of social media and user content aggregators like YouTube and Reddit, it is hard to imagine if not outright inconceivable that there would not be an overwhelming financial incentive to platform user-generated speech and content even in the absence of section 230. In other words, since the internet has already developed a great deal since the passing of the CDA, and there are demonstrable, massive economic benefits to hosting user content, there is no reason to believe that the chilling effect liability could have on content hosting platforms would have a substantial impact on the number of opportunities people have to voice their thoughts online.

Moreover, abuse of such platforms can cause serious and even irreparable harm to people's reputations. Consider the case of Nadire Atas, a Canadian woman and vexatious litigant who serially defamed anyone she disliked. People targeted by Atas (such as her mortgage lender, a family that employed her 30 years ago, lawyers who had represented her, all those people's friends and family, etc.) found articles written about them popping up in google search results

[25] "DEPARTMENT OF JUSTICE'S REVIEW OF SECTION 230 OF THE COMMUNICATIONS DECENCY ACT OF 1996."

accusing them of being pedophiles, scammers, and thieves. These accusations were totally baseless and fallacious, but that didn't stop them from doing serious damage to their victims' reputations. Suppose I was to write an article accusing John Doe of being a pedophile. Even if there is no factual basis for that claim, my defamatory article will still show up whenever someone googles "John Doe", and the mere allegation is often enough to prevent prospective employers from hiring Mr. Doe. This is precisely what Atas did to Guy Babcock and his family, as well as to many other people.[26] This problem is exacerbated by the hands-off approach many tech companies take to content moderation, and their often-slow response to address problems like defamation. In the Atas case, it took upwards of six years of pleading by victims before Google finally moved to delist the defamatory articles. Some sites, like Rip-off Report even monetize the removal of posts, a practice tantamount to extortion. YouTube has seen similar abuse of its copyright protection system, with videos being falsely copyright claimed in order to steal their ad revenue or even just to prevent them from being seen.[27] Most of this detection is automated and it can be difficult to have it overturned even when the content usage falls clearly within fair use guidelines. As all of this behavior is protected under the CDA, there is little incentive for companies to care about what happens to the victims of platform abuse. Removing the liability shield provided by the CDA would incentivize platforms to monitor and address abusive behavior such as Atas'. Liability is thus an especially promising solution because content moderation needs to be platform-specific (as different communities have different rules and tendencies), and the incentive liability provides lends itself to a more proactive approach on the part of hosting platforms to address such abuses.

[26] Hill, "A Vast Web of Vengeance."
[27] Dodgson, "YouTube Channels Are Being Held Hostage with False Copyright Claims, but the Platform's Hands Are Tied."

We could pair these two suggestions (repealing section 230 and concocting a "too-big-to-fail" standard for massive platforms like Facebook and Twitter) to further limit scope and applicability. This would greatly reduce the risk of smaller user content-hosting sites being harassed with lawsuits. We very much want to avoid a scenario in which people are able to bankrupt smaller sites via legal fees, as such tactics would allow anyone with enough money to silence those without the funds to fight them off. Peter Thiel's involvement in the Gawker-Hogan lawsuit and Gawker's subsequent downfall should make us wary of this power.[28] By outsourcing enforcement to private companies via the threat of liability, we can also avoid the need to create an expensive bureaucracy to handle these tasks on its own.

Bear in mind that for a platform to be held liable for harm resulting from user content it hosted, that platform must bear considerable fault or negligence. To help illustrate what I mean by fault or negligence, let me give an example. Suppose I run a YouTube channel which happens to be my sole source of income, and someone files false copyright claims against all my videos, demonetizing my entire channel. I contact YouTube to request a manual review and contest the claims. YouTube might respond promptly and resolve the dispute, but for the sake of this example let us suppose that—as is often the case—they do not. Time here is of the utmost importance, because it matters very much to me whether the issue is resolved within two weeks or two months as it greatly impacts my ability to pay the next month's rent or put food on the table. Were this to happen, and I suffered serious financial harm as a result of YouTube's inaction, I may very well have a credible case against them, regardless of the fact that it was primarily the person who filed the false claims that had wronged me. If there is concern over a flood of trivial lawsuits, then we can simply set a higher bar for harm stemming from platform

[28] Sorkin, "Peter Thiel Is Said to Bankroll Hulk Hogan's Suit Against Gawker."

abuse (e.g. loss of livelihood, risk of assault by hostile strangers, complete destruction of one's reputation, etc.). We could also come up with some sort of "fair use" policy to define slander, which could require a minimum degree of offense, demonstrated malicious intent, evidence of fabrication or baseless accusations, etc. for potential litigation. Otherwise, we can imagine how vexatious litigants could sue over every insult or accusation against their character.

These two measures in tandem would restore some degree of accountability and oversight for the largest web platforms. They would allow those wronged by the negative externalities resulting from abuse of speech on platforms to seek redress for platform negligence whilst simultaneously offering an avenue of recourse for those who had been substantially wronged by the abuse of platform censorship power. Most importantly, perhaps, it would allow us to bring our treatment of speech online more in line with how we treat speech in the "real world," reining in the freedom from consequences currently used by platforms and users alike.

Conclusion

At the heart of the debate over speech lies the ancient questions of what is true, and who gets to decide. The deciding factor in such debates is often the extent to which a thinker is willing to believe in society's ability to "muddle through". Plato feared even the relaxation of aesthetic theater standards, seeing it as the beginning of a downward spiral culminating in lawlessness and ruin. While others since have tended to be more optimistic about toleration, it is only with the rise of liberalism that we see a principled regard for toleration and frameworks of rights, thus giving us free speech. Nevertheless, examination of the historical debate over what speech can be tolerated makes clear that rights always coexist in tension with one another, and the inevitability of rights conflicts entails some necessary limits to their exercise. American legal interpretations of speech emphasize strict scrutiny of any restriction on speech, and principally disallow any prior restraint, content-based discrimination, or political orthodoxy.

As can be seen from the prevalence of restrictive local ordinances, it is commonplace for speech to be restricted in different public contexts. We must also of course bear in mind the role that social approbation and disdain play in regulating our discourse. Reputational counterincentives increase the cost of spreading disinformation and engaging in rude and vitriolic behavior, and traditional publishers act as gatekeepers by vetting would-be publications. The internet introduces new challenges due to the extent to which it provides anonymity, connects people with receptive audiences, and democratizes publication power. While these are typically considered good outcomes, unfortunately it means that the internet is also a prime breeding ground for disinformation. And as the internet has become inextricable from our daily lives and public discourse, we cannot dismiss it as divorced from reality and incapable of

affecting the material world in serious ways. As large social media platforms have become particularly important to this discourse but lack any regulation governing speech, they currently present a large yet unaccountable influence on how we discuss ideas. By selectively repealing section 230 of the CDA, we could introduce liability for the largest platforms. This would allow us to hold them accountable for instances of harm resulting from platform abuse in which the platform is complicit or culpable of negligence, whilst simultaneously creating an incentive for platforms to avoid harm and wanton censorship wherever possible. This could be paired with additional regulatory constraints to check the abuse of platform censorship power and protect some basic degree of speech freedom on the largest and most public online communities of discourse.

Already we have seen myriad examples of the cost to human life and welfare which disinformation can impose upon us, whether in the form of COVID deaths or an attempted insurrection. We therefore have great reason to be concerned over the internet's power to bypass conventional social checks against bad behavior and undermine assumptions implicit in the marketplace of ideas model upon which the rules of our discourse depend. We may not be able to avoid all of the costs of disinformation without unduly restricting individual autonomy of choice, but this does not mean that we should give up on trying to mitigate them. If we are to have any chance at properly addressing and combatting the problems posed by disinformation, we will need to clarify our consensus of interpretation on the foundations and limitations of speech rights, especially within popular discourse. Otherwise, we risk being left behind by the web, unable to cope with the wrenches it throws into our mechanisms of consensus formation.

Acknowledgements

I would like to thank Professor Christopher Freiman of the Philosophy Department for serving as advisor on this project. I would like to thank Professor John Lombardini of the Government Department and Professor Joshua Gert of the Philosophy Department for reading drafts of this thesis and serving on my defense committee. It has been a tremendous learning experience and a pleasure working with them.

I would also like to thank Professor Timothy Zick of the William & Mary Law School for his assistance and advice on legal aspects of this paper.

Special thanks to my father Paul Forte for his tireless support of this project.

Bibliography

Boyer, Pascal. *Minds Make Societies: How Cognition Explains the World Humans Create*, 2018.

Chamberlain, Andrew, Daniel Zhao, and Amanda Stansell. "Progress on the Gender Pay Gap: 2019." Glassdoor, March 27, 2019. https://www.glassdoor.com/research/gender-pay-gap-2019/.

"DEPARTMENT OF JUSTICE'S REVIEW OF SECTION 230 OF THE COMMUNICATIONS DECENCY ACT OF 1996," June 3, 2020. https://www.justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996.

Dodgson, Lindsay. "YouTube Channels Are Being Held Hostage with False Copyright Claims, but the Platform's Hands Are Tied." Insider. Accessed May 9, 2022. https://www.insider.com/youtubers-channels-are-being-held-hostage-with-fake-copyright-claims-2020-6.

Feinberg, Joel. *The Moral Limits of Criminal Law: Offense to Others*. Vol. 2. New York: Oxford Univ. Pr, 1987.

Fish, Stanley Eugene. *There's No Such Thing as Free Speech, and It's a Good Thing, Too*. New York: Oxford University Press, 1994.

Grendler, Paul F. *The Roman Inquisition and the Venetian Press, 1540-1605*. Princeton, N.J: Princeton University Press, 1977.

Hill, Kashmir. "A Vast Web of Vengeance." *The New York Times*, January 30, 2021, sec. Technology. https://www.nytimes.com/2021/01/30/technology/change-my-google-results.html.

Kay, Grace. “The Many Alleged Identities of Bitcoin’s Mysterious Creator, Satoshi Nakamoto.” Business Insider. Accessed May 9, 2022. https://www.businessinsider.com/bitcoin-history-cryptocurrency-satoshi-nakamoto-2017-12.

Kodan, Shaivya. “🧟 Zombie Statistics.” *Medium* (blog), March 27, 2021. https://shaivyakodan.medium.com/%EF%B8%8F-zombie-statistics-97247986c6e8.

Kolbert, Elizabeth. “Why Facts Don’t Change Our Minds.” *The New Yorker*, February 19, 2017. http://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds.

Locke, John. *Second Treatise Of Government*. Accessed May 9, 2022. https://www.gutenberg.org/files/7370/7370-h/7370-h.htm.

Mill, David van. “Freedom of Speech.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/spr2021/entries/freedom-speech/.

Mill, John Stuart. *On Liberty*. Accessed May 9, 2022. https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/mill/liberty.pdf.

Milton. *Areopagitica*. Accessed May 9, 2022. https://milton.host.dartmouth.edu/reading_room/areopagitica/text.html.

Plato. *Complete Works*. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis, Ind: Hackett Pub, 1997.

Roose, Kevin, and Mike Isaac. “Facebook Dials Down the Politics for Users.” *The New York Times*, February 10, 2021, sec. Technology. https://www.nytimes.com/2021/02/10/technology/facebook-reduces-politics-feeds.html.

Shank, J.B. “Voltaire.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University, 2021. https://plato.stanford.edu/archives/sum2021/entries/voltaire/.

Sørensen, Anders Dahl. *Plato on Democracy and Political Technē*. Philosophia Antiqua, volume 143. Leiden Boston: Brill, 2016.

Sorkin, Andrew Ross. “Peter Thiel Is Said to Bankroll Hulk Hogan’s Suit Against Gawker.” *The New York Times*, May 25, 2016, sec. Business. https://www.nytimes.com/2016/05/25/business/dealbook/peter-thiel-is-said-to-bankroll-hulk-hogans-suit-against-gawker.html.

Spinoza, Benedictus de. *Theological-Political Treatise*. Edited by Jonathan Israel. Cambridge Texts in the History of Philosophy. Cambridge ; New York: Cambridge University Press, 2007.

Warzel, Charlie. “Is QAnon the Most Dangerous Conspiracy Theory of the 21st Century?” *The New York Times*, August 4, 2020, sec. Opinion. https://www.nytimes.com/2020/08/04/opinion/qanon-conspiracy-theory-arg.html.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First Trade Paperback Edition. New York: PublicAffairs, 2020.