5-2022

A Dangerous Pursuit of Safety: The Value of Freedom of Expression and the Folly of Trading It Away

Nina Katarina Štular
DePauw University

A Dangerous Pursuit of Safety: The Value of Freedom of Expression and the Folly of Trading It Away

*
**

Nina Katarina Štular
DePauw University Honor Scholar Program
Class of 2022

Sponsored by Dr. Kevin Moore
First Reader: Dr. Marcia McKelligan
Second Reader: Dr. Michael Seaman

Acknowledgements

I would like to thank my sponsor Kevin whose expertise and insight helped me draw connections between disciplines and whose enthusiasm to discuss these topics with me kept me motivated throughout the year. My deepest gratitude also goes to my advisor and friend Marcia McKelligan, without whose mentorship, guidance, and unwavering support throughout my time at DePauw I would have neither skill nor confidence to tackle such a topic. I wish to thank Prof. Michael Seaman for furthering my understanding of free speech in Ancient Greece and Prof. Ted Bitner for sharing with me a clinical psychologist's perspective through conversations which I found both a great source of inspiration for this work and personally enriching. Lastly, I thank Christian for helping me see the value of taking breaks and for patiently reminding me not to overthink everything.

Table of Contents

Introduction

"Tis slavery, not to speak one's thought"

-Euripides, *Trojan Women*

People love freedom of expression in the abstract but not in practice. In a survey conducted across 33 countries the median percentage of participants in support of the statement "people can say what they want" was 94 (Skaaning and Krishnarajan, 7). When asked whether they support the freedom to express something offensive to their religion and belief, however, the same participants drop their support to 61 % and further lower it to 41% when it comes to expression offensive to minority groups (Skaaning and Krishnarajan, 8). It is only natural for people to want to censor those spreading the ideas they disagree with, dislike, and fear. Expression offends and upsets, it can lead to emotional pain, and bring about suffering. It allows those with malicious intentions to organize, garner support, and execute horrific deeds. Expression is a tool of deception and manipulation. There are plenty of easily accessible reasons for people to hate it, fear it, and see it fit to persecute and censor it.

For the majority of history, the burden of proof rested on the shoulders of those few individuals who viewed free expression as valuable. Unsurprisingly, institutions of power have been quite comfortable with the idea that "dangerous" expression must be censored for the common good. Especially at moments when a new technology made expression more accessible to people, authorities time and time again eagerly stepped up to prevent the everyman from corrupting the masses through the spread of unvetted ideas. The arguments warning against the dangers of misinformation on the internet today are a copy of those used at the dawn of the printing press, the telegraph, the newspaper, and the radio (Mchangama, 5). There is a trade off at play, we are told by those in power: we can have freedom of

expression, or we can have safety. While the former may have an abstract appeal, safety is what anyone will choose in practice.

In the chapters ahead I aim to explore the relationship between free expression and safety and put the supposition that they are in conflict under question. In the first part I will explore reasons for valuing free expression that come to light in the examination of history of philosophy. By presenting why Ancient Athenians valued free speech, Locke advocated for religious toleration, and Mill defended non-conformism, I hope to provide a strong case for viewing free expression as a fundamental human liberty integral to our personal and public well-being. I will conclude the section by applying evolutionary principles to ideas in order to provide an alternative argument for the invaluable role free expression plays in the construction of knowledge across generations.

In the second part I will explore the reasons behind restricting freedom of expression. There are at least some instances in which concerns for public safety indeed override the importance of permitting expression, but there is widespread disagreement about where the pool of permissible restriction ends. After a brief overview of the current international, US, and EU legislation, I will forward three standards that ought to govern speech regulation: The Severity Standard, The Direct Causality Standard, and The Last Resort Standard. I will apply the three standards to three most common grounds upon which expression is regulated: harm, offense, and hate. I hope to show that the only cases in which expression should be prohibited or penalized by a governmental authority are those in which censorship is the only way to avoid a severe harm that expression would directly cause. I will conclude the section by discussing the way findings from science are deceptively employed to garner support for less narrow speech regulation. I hope to show that even though many want them to, neuroscience

and psychology do not show that expression regulation is the only, or even the best way of tackling group divisions, hatred, and violence.

In part three I will identify the idea that speech equals violence as the key assumption backing many contemporary pro-censorial attitudes. There is a long and well discussed historical tradition of censoring expression to protect respect for a governmental or religious authority or the moral purity of a community. I chose to focus on censoring expression for the sake of safety, however, because of the fast growing popularity of pro-censorship arguments appealing to safety in our public sphere and especially academia. I will argue that the desire to restrict expression to create "safe environments" stems from a shift in discussing language as a literal tool of aggression and violence–while language undeniably has power, its power ought not be understood as equal to that of a punch. After presenting the ways in which some kinds of speech may be metaphorically portrayed as violent, I will show that understanding language as an opposite to violence is integral to our ability to morally condemn violent behavior. I will conclude the section by examining how findings from psychology and neuroscience apply to the question of whether speech equals violence.

Part four will consist of an exploration of the ways in which playing into the "speech=violence" fallacy affects educational and academic environments. Because the shift in understanding language as literally harmful is in its origins academic, the culture forming in the classrooms and college campuses can serve as a warning against letting the oversensitivity seep into the public sphere. After discussing the hasty and misguided popularization of trigger warnings in classrooms, I will analyze the culture of speech-sensitivity created on US college campuses by those who understand safety as incompatible with free expression. I will argue that through glorifying emotional safety, academic institutions have paradoxically become a dangerous place for intellectual activity of

their students. Finally, I will look at three highly publicized smear campaigns activists ran against reputed academics because they assumed their work to be a threat to the safety of vulnerable groups. Such case-studies will serve as examples how quickly the desire to protect can propagate injustice if guided by emotional thinking rather than a cautious, rational approach to activism.

Freedom of expression has always been weighed against other values. Like all freedoms, it brings with it a degree of risk that it will be misused to do ill. In the work ahead, I hope to offer a reminder of the value of free expression and a warning against the folly of over-restricting it. Those in favor of censoring and regulating expression too often feign to act as protectors of the much desired public or community safety. What censorship really produces, however, is at best a mere illusion of safety. Only an environment in which all ideas, no matter how unpleasant, can be debated, challenged, and questioned is one in which human liberty is respected and social problems can be diagnosed and addressed before they fester into action.

PART 1: WHY FREEDOM OF EXPRESSION MATTERS

A. The Historical and Philosophical Background of Freedom of Expression

1. Ancient Athens

Freedom of speech became an important value as early as in the 5th century BC in the Greek city state of Athens. Its origins are inseparable with the development of Athenian democracy so much so that it is difficult to say whether democracy gave birth to freedom of speech or whether freedom of speech gave birth to democracy. In Ancient Greece, democracy was a political system based on liberty or *euletheria* of which freedom of speech was the "most important and necessary ingredient" (Momigliano, 259). The strength of this connection relied on the fact that the Assembly—the main decision-making body in Ancient Athens that consisted of all law-abiding male citizens—granted its members the liberty to participate in the political decision-making of the polis precisely by addressing each other and commencing sessions with the famous call: "Who wishes to speak?" (Hansen, 89) In a short treatise on the Athenian constitution known as The Old Oligarch, its author famously highlights this link between democracy, freedom of speech, and self-deliberation as he complains that in a democracy "any worthless person who wishes can stand up in the assembly and procure what is good for himself and those like him" (Xenophon, 34). Indeed, freedom of speech sets democracy apart from other authoritarian systems as remarked by the orator Demosthenes: "a basic difference between Spartan oligarchy and Athenian democracy is that in Athens you are free to praise the Spartan constitution and way of life, if you so wish, whereas in Sparta it is prohibited to praise any other constitution than the Spartan" (Hansen, 77). To expose the nature of the Greek conception of freedom of speech, I will first examine two core concepts: *isegoria* or equality in speech and *parrhesia* or unbridled speech.

I will continue by laying out three reasons for which freedom of speech is valued in Ancient Athens and conclude by discussing its limitations.

I. Isegoria

The first concept at the core of Athenian freedom of speech is *isegoria* or the ability of every free citizen to address the assembly. This equality of speech is such an essential trait of democracy that Herodotus uses it to signify the democratic type of governing (Hansen, 83; Momigliano, 259). However, *isegoria* establishes equality of opportunity and does not include natural equality or equality of outcome: "No Athenian expected that every one of the 6000 citizens who attended a meeting of the Assembly could - or would - address his fellow citizens. *Isegoria* was not for everyone, but for anyone who cared to exercise his political rights" (Hansen, 83). Exercising your freedom to speak in Athens, however, required a great deal of courage: "speaking before one's fellow citizens required the ability to take risks, to confront social dangers such as humiliation or shame, and to maintain a level-headed composure in expressing one's own opinions before one's opinionated political equals" (Balot, 259). Not only humiliation and shame, but sometimes even graver dangers threatened the speaker because of *thorubos*, namely the phenomenon of communal jeering, yelling, and being altogether raucous through which the *demos* expressed its opinion of what was being said: "through the *thorubos*, the Athenians made it clear, through either interrupting or not, that the *demos* was all-powerful and was ultimately responsible for the speakers' freedom to speak altogether" (Balot, 258). Rather than an inalienable right that protects the individual from the government, Athenians saw equal opportunity to speak as well as speaking frankly as a trait of their system.

II. Parrhesia

Parrhesia or the ability of citizens to speak frankly like *isegoria* ought not be viewed as a negative right: "what the Athenians called *parrhesia* was more a characteristic of their citizenship than it was a right" (Carter, 211). The unbridled, uninhibited, honest speech, sometimes referred to as "a citizen attribute" (Carter, 229) was thus a non-ideological trait of those whom the political system encourages and enables to speak their mind rather than taper their expression to please an oligarch or a monarch. In Athens *parrhesia* is associated with honesty and simplicity of expression and is thus seen at odds with the art of rhetoric that teaches the speaker how to use language to persuade and manipulate rather than simply express: "*Parrhesia* opposed rather than supported the practice of a rhetoric that obscures and distorts the truth for the sake of individual benefit" (Saxonhouse, 104). Demosthenes highlights this opposition by concluding his *Philippic IV* with a declaration of his commitment to *parrhesia*: "There you have the truth spoken with all freedom, simply in goodwill and for the best—no speech packed by flattery with mischief and deceit and intended to put money into the speaker's pocket and the control of the State into our enemies' hands" (10.76). Athenians therefore view the principle of *parrhesia* as something that promotes the interests of the community, often at a personal risk of shame and humiliation. However, what was the perceived value of nourishing such a culture of free expression based on equality of opportunity, frankness, and honesty?

III. Self-determination

Herodotus in *Histories* connects *isegoria* with the ability to self-govern and determine one's own political fate but also with an increased initiative of individuals to participate in politics. In book 5 he writes: "So the Athenians grew in power and proved, not in one respect only but in all, that equality [in speech: *isegorie*] is a good thing. Evidence for this is the fact

that while they were under tyrannical rulers, the Athenians were no better in war than any of their neighbors, yet once they got rid of their tyrants, they were by far the best of all. This, then, shows that while they were oppressed, they were, as men working for a master, cowardly, but when they were freed, each one was eager to achieve for himself" (5.78). He presents what can be seen as an early version of the later libertarian idea that leaving people free to pursue their own perhaps at times selfish interests leads to public good. Freedom of speech that grants the Athenians the ability to participate in the political-decision making, according to Herodotus, leads to greater military success, perhaps precisely because one fights with greater fervor for one's own cause than for that of a tyrant.

IV. Decision-making Accuracy

Another reason for which the Athenians value free speech is because it leads to a greater quality of decision-making. Demosthenes in *Exordia* emphasizes that the freedom of Athenians to raise objections and concerns about the proposed course of action guards the city from making rash decisions: "In my opinion, men of Athens, no intelligent citizen would deny that it is best of all for the city, preferably at the outset not to do anything inexpedient, but otherwise, that those should be on hand who will object at once" (Ex. 49. 1). In other words, Demosthenes emphasizes that a decision maker more likely succumbs to emotion or folly and makes a poorer decision if no one dares to disagree with him. In order for Athens to reap the benefits of communal decision-making, however, Demosthenes sees it crucial that people listen to each other: "To this must be added, however, that you shall be willing to listen and learn; for nothing is gained by having a man who will give the best counsel unless he shall have people who will listen to him" (Ex. 49. 1). Through this subtle critique of *thorubos*, Demosthenes emphasizes the risk of the Athenians to self-sabotage and undo the societal benefits of free speech by too rapidly suppressing displeasing speech.

V. True Courage

Apart from enabling self-deliberation and bettering the quality of decision-making, Pericles in the famous funeral oration from Thucydides's *History of the Peloponnesian War* links freedom of speech to the ability to develop true courage. Because Athenians often needed to defend their political systems against accusations of military inefficacy, Pericles considers it important to first emphasize that deliberation does not impede action: "The great impediment to action is, in our opinion, not discussion, but the want of that knowledge which is gained by discussion preparatory to action" (42. 2). He continues by drawing a distinction between being brave due to ignorance and being courageous as a result of having carefully considered a decision before acting: "For we have a peculiar power of thinking before we act and of acting too, whereas other men are courageous from ignorance but hesitate upon reflection. And they are surely to be esteemed the bravest spirits who, having the clearest sense both of the pains and pleasures of life, do not on that account shrink from danger" (42.3). By juxtaposing the ability to discuss a decision with the ability to think it through, he links the two to informed and decisive action which he presents as the basis for exercising courage.

VI. Limitations

The value of freedom of speech is instrumental for the Athenians. They adopt it because they see democratic and discussion-based decision-making as means to a greater military and political success. Free speech belongs primarily to the Assembly and is subject to the will of the *demos*. The most famous case in which the *demos* punishes speech is the trial of Socrates in 399 BCE at which the famous philosopher receives a death sentence for not believing in the gods the city believed in and for corrupting the youth. Some consider it a great failure of free speech and deliberation that the jury found Socrates guilty. Others,

however, emphasize that in no place apart from Athens, could Socrates have been so free to openly critique the city's way of governing for decades without punishment. If we view Socrates as a public threat, it could possibly be argued that Athenians punished speech in only the cases of the most severe offense and "limited free speech (and other freedoms) only to protect the democracy from substantive, material harm" (Wallace, 213). Hansen, however, points out that Plato and Aristotle write most unfavorably about democracy as well which he sees as grounds to understand Socrates's trial as "an isolated occurrence" in what is otherwise an admirable legacy of free speech in Athens: "the criticism of democracy to be heard in Athenian sources is the strongest possible evidence that the Athenians' pride in their freedom of speech was not unfounded" (26). Regardless of which interpretation we sympathize with, Socrates's execution happens because the Athenians do not understand free speech as something of intrinsic worth that belongs to every individual. Rather, they view it as a public tool with instrumental value that may be overridden by greater concerns such as presented by an old man questioning the political and religious backbone of Athenian society.

2. The Early Modern Period

One of the most important steps in the historical development of the value of free expression as conceived by modernity is the transition from viewing free expression as an instrumental good subject to being overridden versus regarding it as a fundamental inalienable right of intrinsic worth to every human being. The ideas of John Locke, an English 17th century philosopher, most notably helped facilitate the leap. James Madison, the drafter of the most important piece of legislation protecting the right of expression, "adapted Lockean principles to defend freedom of speech" (McGinnis, 60) in the First Amendment. To readers of Locke it should come as no surprise that the founding fathers of the US would find

principles around which to build a liberty-preserving society in Locke's writings. After all, in his writing Locke notably defends the right to self-determination by arguing that people are "by nature all free, equal, and independent" (49) and that as such they ought to be able to "dispose of their possessions and persons as they think fit" (4). Because the government's power according to Locke comes from people's free decision to unite, a government may exercise political force over its people only to ensure public safety and protect fundamental liberties of the individual. Accepting Locke's conclusion that there are certain inalienable rights deserving of governmental protection, however, does not in itself establish that freedom of speech is one of them. To examine the important transition that leads to the inclusion of freedom of speech amongst a person's most basic rights, I will take a close look at Locke's arguments in *A Letter Concerning Toleration* and discuss their implications. I aim to show that Locke's philosophical works set the ground for elevating freedom of speech from something instrumentally useful to an intrinsically valuable right—an idea that flourishes in the writing of the First Amendment.

I. Religious Toleration in Context

In 17th century Europe religious toleration was on shaky grounds. On the continent, the Peace of Augsburg (1555) which ensured that the ruler could choose the religion for his subjects (*cuius regio, eius religio*) collapsed at the beginning of the century and began the Thirty Years' War in the Holy Roman Empire. While the Jacobean era in England brought some relief from religious strife, "Bloody Mary's" burnings of religious dissenters were far from forgotten. Because of the tense political circumstances Locke himself fled England and situated himself in the more religiously tolerant Dutch republic in 1683 where he wrote *A Letter Concerning Toleration*. His call for toleration and secularism came after centuries of religious discord during which the common attitude was religious uniformity, namely the

practice of the government to promote one religion at the cost of other beliefs.[1] Religious identity at the time, however, greatly defined all areas of an average person's life as religious rituals and holidays made up a big part of a community's social life. Moreover, an intellectual's education included interpretation of sacred texts and important philosophical deliberations of the time often revolved around questions of religious dogma. Enforcing a certain religion therefore significantly affected people's daily lives and put limits to their intellectual activities

Given that such a political climate makes severe restrictions of people's freedoms quite commonplace, Locke's defense of religious toleration aptly focuses on re-establishing individual's liberty as the capital value of any just society. William Popple, a Unitarian merchant and religious writer states the diagnosis of the political situation in 17th century Europe presumed to motivate Locke's writing as lack of liberty: "Absolute Liberty, Just and True Liberty, Equal and Impartial Liberty, is the thing that we stand in need of" (123). In the letter that follows, Locke voices his defense of freedom of religion, ultimately arriving at the conclusion that "Liberty of Conscience is every man's natural Right" and that "no body ought to be compelled in matters of Religion, either by Law or Force" (159). He supports his argument by two separate lines of reasoning, one that outlines the nature of the government and the church, and the other that focuses on the nature of belief. He argues that the government should not and cannot curtail an individual's freedom of belief.

II. Nature of the Government and the Church

The first way Locke goes about defending religious toleration is by arguing that it is neither the business of the government, not the church to force religion onto people: the government ought not interfere in spiritual matters that determine the soul's salvation and

[1] For a more detailed discussion of historical circumstance see Murphy's *Conscience and Community: Revisiting Toleration and Religious Dissent in Early Modern England and America*

eternal life, whereas the church ought not meddle in laws or employ civic force. He begins by defining the government as "a Society of Men [sic] constituted only for the procuring, preserving, and advancing of their own Civil Interests" and defines civil interests as individual's person and property: "Civil Interests I call Life, Liberty, Health, and Indolency of body; and the Possession of outward things, such as Money, Lands, Houses, Furniture, and the like" (128). He sees the government as overstepping its role in interfering with an individual's religion because he views religious beliefs as personal and an individual's freedom to live according to them as crucial to self-determination: "it appears not that God has ever given any such Authority to one Man over another, as to compell any one to his Religion. Nor can any such Power be vested in the Magistrate by the Consent of the People; because no man can so far abandon the care of his own Salvation, as blindly to leave it to the choice of any other, whether Prince or Subject, to prescribe to him what Faith or Worship he shall embrace" (128). In other words, the choice of religion is too important to be outsourced. Because in choosing to follow a religion, a believer works towards ensuring afterlife for his soul, one must choose one's faith alone and carry the responsibility of one's choice. The government may regulate public affairs by, for example, running healthcare and training the military, but it has no claim to people's spiritual life.

Unlike the government, the church concerns itself with people's spiritual life. Because the primary goal of the church is to save souls, Locke argues that it should not use any form of worldly power to coerce people to become its members. According to Locke the church is "a voluntary Society of Men, joining themselves together of their own accord, in order to the publick worshipping of God, in such a manner as they judge acceptable to him and effectual to the Salvation of their Souls" (130). He sees great importance in emphasizing that just as members of church join voluntarily, they may voluntarily leave if unconvinced by the

doctrine: “No Member of a Religious Society can be tied with any other Bonds but what proceed from the certain expectation of eternal Life” (130). By appealing to Matthew 8:20,[2] he refutes the objection that a religion needs authority figures such as Bishops or apostles who need to be able to exercise some sort of power. He as effectively addresses the objection that power is necessary for enforcing religion’s laws by emphasizing that forcing people to abide by certain religious rules is useless if they are internally disinclined to follow them (130). To highlight that it is necessary for religious institutions to refrain from exercising force, Locke points at the situation familiar to his reader: the chaos of religious strife that originates in religions’ conviction that their own doctrine is correct and that they may exercise power to forward it. He asks: “But if one of those Churches hath this Power of treating the other ill, I ask which of them it is to whom the Power belongs, and by what right?” (135). Many have claimed to have the one true answer to this question, but each argues in support of their own religion. Locke’s remark that “every Church is Orthodox to itself; to others, Erroneous or Heretical” (135) thus bears much truth and highlights the importance of tolerance for peace. If both the government and the church were strictly fulfilling their functions, tolerance would be a given. Yet due to their tendency to overstep their jurisdictions, religious conflicts remain a sad reality to this day.

III. Nature of Belief

The second line of reasoning Locke adopts in support of toleration consists of highlighting that personal and internal nature of belief renders any use of force futile. He emphasizes that true significance of religion lies in its ability to affect people’s minds: “All the Life and Power of true Religion consist in the inward and full perswasion of the mind” (128). Therefore, demanding external compliance with rules of religion is counterproductive

[2] In Matthew 8:20 Jesus says: “For where two or three gather in my name, there am I with them.”

to religion's aim to save souls from damnation: "Whatever Profession we make, to whatever outward Worship we conform, if we are not fully satisfied in our mind that the one is true… such Practice, far from being any furtherance, are indeed great Obstacles to our Salvation" (128). Because thought and action differ in natures, changing people's actions by force does little to reform their privately held beliefs: "And such is the nature of the Understanding, that it cannot be compell'd to the belief of any thing by outward Force. Confiscation of Estate, Imprisonment, Torments, nothing of that Nature can have any such Efficacy as to make Men change the inward Judgment that they have framed of things" (129). Locke thus proposes that rather than by force, minds can only be changed, and religion effectively spread by non-violent rational means such as discussion and reasoning: "It is only Light and Evidence that can work a change in Men's Opinions. And that Light can in no manner proceed from corporal Sufferings, or any other outward Penalties" (129). He illustrates his theory of the nature of belief-reforms by personifying truth as an entity which carries enough power and significance to impress itself onto people's mind without the aid of external force: "if Truth makes not her way into the Understanding by her own Light, she will be but the weaker for any borrowed force Violence can add to her" (153). Discussing truth in such terms may be an idealization, but an attractive claim lies at the core of his argument—the pen is mightier than the sword.

IV. The Value of Free Expression

Even though Locke's arguments advocate for religious toleration, they have been recognized as highly relevant to debates on freedom of expression. Locke's conception of the government's and the church's place in the society and their duties towards an individual establish a strong case for freedom of thought, also referred to as freedom of belief or "Liberty of Conscience." While the matter certainly warrants a more careful discussion, I

hope that for the time being we can acknowledge the close bond between freedom of thought and freedom of speech: being free to think what you want means little without being able to express one's thoughts and vice versa. Those interpreting Locke's writing picked up on the connection between freedom to think and freedom to express one's thoughts and have used it as grounds for establishing freedom of expression as intrinsically valuable.

VI. Property-Based Understanding of Freedom of Speech

Even though Locke never states it explicitly, his interpreters take his philosophy to contain all the necessary elements for believing that Locke would stand against governmental restriction of expression because he would have counted freedom of speech amongst individual's fundamental and inalienable rights. In other words, not only would Locke see instrumental value in freedom of expression, but he would also consider it a member of the natural rights of life, liberty, and property. In counting freedom of speech as a natural right, Locke fundamentally shifts the discussion regarding liberty of expression—free speech that was thus far valued for promoting important societal goods must now be considered as an intrinsically valuable necessary aspect of an individual's dignity.

The interpretation that puts Locke at the forefronts of freedom of speech finds its textual support in Locke's discussion of property. In "The Once and Future Property-Based Vision of the First Amendment," John McGinnis presents Madison's interpretation of Lockean principles and argues that Locke paves the way for freedom of speech to be considered as an important aspect of a person's right to own and manage his own property. McGinnis refers to two of Locke's statements, namely that "every Man has a *Property* in his own *Person*" and that "the *Labour* of his Body and the *Work* of his Hands…are properly his" (15). Madison makes the crucial connection between these statements by explicitly referring to the ideas that a man produces his property as he asserts that a man has "a property right in

his opinions" (66). Madison thus proposes that just as people have a natural right to own and manage property, they have a natural right to own and manage the fruits of their intellectual labors, namely their thoughts and opinions. McGinnis demonstrates the presence of all the ingredients necessary for this conclusion in Locke's writings by bringing up Locke's understanding of property in relation to autonomy. According to McGinnis, Locke holds that an investment of energy into matter grounds an individual's claim to ownership: "an individual owns property by applying his labour to matter and thus infusing his spirit into nature" (66). Just as we work to attain material goods, we invest labor and energy into the production of our own thoughts and therefore have a claim to express them freely. The premises Locke lays down therefore enable Madison to justify granting freedom of speech such an important place in American legislation.

Because Locke sees the protection of private property as one of the key tasks of government, conceptualizing speech as property guarantees his opposition to regulation of freedom of speech. He would, however, oppose speech regulation even without such a move. At the core of Locke's political philosophy lies the idea that government exists to ensure public peace and security and ought to interfere in an individual's life as little as possible: "Any exercise of political power over individual behavior which did not threaten peace or security was an exercise of power unjustified by the end for which that power existed" (Donne, 32). Therefore, according to the spirit of Lockean arguments freedom of speech belongs to basic human liberties as it protects the autonomy of an individual from governmental interference. Locke presents thought and expression as fruits of our labor in a way that makes it an easy step for Madison to view them as our intellectual property and write that "a man has a property in his opinions and the free communication of them" and that a just government "impartially secures to every man, whatever is his own" (65).

Therefore, even though Locke primarily writes about the freedom of religious belief, his political philosophy presents freedom of thought and expression as an intrinsically valuable trait of a free society and inspires the inclusion of freedom of speech into the American Constitution.

3. Mill

John Stuart Mill, arguably the most famous defender of free expression, offers a comprehensive account of the values of free expression that integrates both intrinsic and instrumental reasons. In *On Liberty* (*1859)* he presents a theory aimed to "make the fitting adjustment between individual independence and social control" (76), namely, to forward an argument as to what laws ought to govern the relationship between the society and the individual. While he declares that he will "forego any advantage" he may get by presenting freedom of expression as intrinsically valuable in the Lockean sense, his arguments bring to light both instrumental and intrinsic value of free speech. Namely, by presenting intellectual liberties of thought, expression, and discussion as means to a healthy society and individual autonomy, Mill constructs such intellectual liberties as the essential attribute of public and private good. Furthermore, he emphasizes that being exposed to all kinds of opinions is a prerequisite for anyone's intellectual flourishment and an effective method of truth-discernment. Behind such seemingly instrumental arguments, however, lies the foundational idea that an intellectually thriving society with autonomous individuals is a good in itself. Mill lays this foundation by arguing that freedom of thought and expression are inseparable; despite its public nature, speech is a self-regarding act whose free exercise is essential for personal autonomy. After discussing Mill's foundational argument for free

speech, I will lay out his three epistemic arguments for free expression grounded in the instrumental importance of free discussion for the pursuit of truth.

I. The Foundation: Free Expression and Intellectual Autonomy

At the very beginning of *On Liberty*, Mill outlines two aspects of human life in which an individual ought to be free: thoughts and actions. While there are instrumental reasons to ensure that an individual may act and think freely—e.g. greater creativity and productivity, Mill does not draw on them. The lack of emphasis on the instrumental good of free thought and action suggests that Mill's reason for bringing up the two basic components of self-determination is to outline the areas of being in which a person ought to be free because individual liberty is an intrinsic good. According to Mill, freedom of action equals a freedom to live as we please or the liberty "of framing the plan of our life to suit our own character; of doing as we like, subject to such consequences as may follow" (83). Such a liberty is relatively straightforward and limited only by the duty to refrain from impeding on the ability of others to act freely as well. Freedom of thought, however, refers not merely to introspective activities such as "liberty of conscience…liberty of thought and feeling…freedom of opinion and sentiment on all subjects" (82). Rather, Mill takes a crucial step in arguing that liberty of expression, namely freedom of speech and publication is an integral part of freedom of thought. By including it under freedom of thought, Mill brings free expression into the realm of intrinsically valuable liberties.

Mill acknowledges that freedom of expression appears other-regarding and as such may not appear an intrinsically valuable liberty, essential to the protection of an individual's dignity. Rather, it appears to be a liberty that we value instrumentally for its social value. Freedom of expression "may seem to fall under a different principle, since it belongs to that part of the conduct of an individual which concerns other people" (82). However, Mill

maintains that thought and expression are “practically inseparable” (83). He states that the two are “resting in great part on the same reasons” (83) thereby implying that one cannot exist without the other. Indeed, that expression depends on thought has immediate intuitive allure: while we colloquially say that people sometimes ‘speak without thinking’ or ‘blurt speech out,’ such instances of quick and spontaneous speech in most cases reveals inner thoughts and sentiments of the speaker rather than attests to their absence. Before expressing any idea, one must, however briefly, first conjure it up in thought. The dependence of thought on expression, however, seems less straightforward. We are certainly capable of having thoughts without expressing them. However, most thoughts that we have in some way or other depend on expression we have witnessed: a complex thought such as ‘capitalism is a superior system to communism’ surely arises in my mind in part because of books that I have read and even a simple thought such as ‘I see a tree’ depends on other people teaching me the meaning of words I need to formulate it. In short, we are limited in our inventory of ideas if we do not have the access to the marketplace in which they are freely expressed. As Rourke in his commentary of Mill suggests, we cannot intellectually flourish without accessing the ideas of others:

> “Without intellectual independence, which can be achieved primarily through education and by encouraging individuals to think for themselves, people cannot be said to be in a position to make an informed choice concerning what constitutes their own good. But people are unable to make an informed choice unless they can consider all sides of a question: freedom of thought is of little or no value if people are not at liberty to hear the private opinions of others. To be deprived of the opportunity to compare and contrast ideas is to be deprived of education and the opportunity to expand one’s intellect.” (78)

In all areas of social life from art and the business world to academia, we cherish originality of thought. However, we often fail to acknowledge that already present ideas enable the birth of what we consider novelty. For example, the literary critic Harold Bloom argues in *Anxiety of Influence* that there is no such thing as a completely original poem—each author steps into a conversation with his predecessors and in his work reworks and reinterprets theirs. Indeed, originality requires inspiration; genius consists of recombining the ideas of the past in new ways. Thus, in order to have a rich intellectual life with diverse and far-reaching thoughts, we must have the thoughts of others at our disposal. We are social animals to the extent that our innermost thoughts and with them the essence of our individuality depends on our fellow humans. Our ability to exercise intellectual autonomy thus depends on the freedom to think and express thoughts to the same extent that our personal autonomy depends on being able to act and move around freely.

Mill implicitly picks up on the intellectual vulnerability of humankind to social pressures when he warns against the influence of the oppression of public opinion. He views the tyranny of majority that uses methods such as social stigmatization and shaming as even more dangerous to intellectual freedoms than political oppression by virtue of legal restrictions—the former "leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself" (76). He defends the right and value of dissent against the "tyranny of prevailing opinion and feeling" (76) and warns that societal norms tend to "prevent the formation, of any individuality not in harmony with its ways" (76). Once again, the implicit reasons behind Mill's warnings lean on the assertion that there is intrinsic value in a liberated, self-determining mode of being. Mill could have brought up instrumental reasons for preserving difference of opinion: history offers us plenty of instances when the majority strayed in their judgment of right and wrong and non-conformism saved

the day. To name just few examples, loudly stated dissent from majority opinion served as a spark for Reformation, female suffrage, and The Civil Rights Act. While historical examples present the instrumental value of non-conformism, Mill aims to ground his defense of dissent on opinion in a more fundamental assertion, namely that "there is a limit to the legitimate interference of collective opinion with individual independence" (76). Mill's defense of free expression and his emphasis on the need to protect an individual's right to dissent relies on a basic idea that protecting an individual's intellectual autonomy is a good in itself that ought not to be overridden by the inconvenience caused by the dissemination of offensive, outrageous, and unpleasant ideas.

II. The Utility of Free Expression for Truth-Seeking

Even though Mill hints at intrinsic reasons for valuing free expression in the introductory arguments of *On Liberty*, the most famous part of Mill's writing consists of epistemological arguments. In three separate arguments, Mill outlines the instrumental value of free expression—it contributes to the pursuit of truth. First, Mill argues that stifling an opinion falsely assumes infallibility, second, he argues that even false opinions contribute to truth, and third, he argues that many beliefs contain partial truths. His arguments establish that there should be no interference with the discussion of any opinion, however unpleasant and controversial.

a) Infallibility

Mill argues that whoever wishes to stifle an opinion illegitimately asserts that their own view is absolutely certain: "All silencing of discussion is an assumption of infallibility" (88). He sees the restriction of other people's opinions on the assumption of infallibility as an epistemic flaw. No one has grounds to be absolutely certain in their own right because of the indirect nature of experience—we do not interact directly with the world,

but merely interpret our subjective experience by engaging imperfect faculties of judgment. Therefore, absolute certainty is unattainable. On such a simple interpretation, Mill's argument opens itself up to the kind of objections that plague simple truth relativism that claims we cannot know anything to be true: if we cannot assume infallibility about any opinion, then we cannot assume that this very ban of assuming infallibility is infallible and applies universally. However, Mill's argument does not aim to prohibit an individual from forming beliefs and acting according to them due to the lack of absolute certainty. Mill recognizes that there is "assurance sufficient for the purposes of human life" that allows us to act according to our beliefs and to "assume our opinion to be true for the guidance of our own conduct" (89). Because we cannot have absolute certainty, however, marking someone else's belief as absolutely false and prohibiting its discussion assumes one's own opinion as absolute truth. Mill emphasizes that personally we may hold a conviction and advocate for it without closing off the possibility of it being disputed and overruled by new evidence: "There is greatest difference between presuming an opinion to be true, because, with every opportunity for contesting it, it has not been refuted, and assuming its truth for the purpose of not permitting its refutation" (89). New experience and data help us correct our misconceptions, but we need discussion to interpret and bring meaning to them (90). We need to remain open to the possibility that we are wrong because assuming infallibility is epistemically dishonest.

Besides expressing epistemic concerns, however, Mill condemns assuming infallibility by imposing our own opinions onto others as a moral violation of our responsibility to allow people to form their own opinions. Mill argues that to prohibit someone else's opinion, our standard of certainty must be much higher than the level of certainty that we need to make decisions on our own. In fact, to force others to hold an opinion, we would have to know it to be true with absolute certainty—anything less than

certainty unfairly imposes onto others our own epistemic risk of being wrong. Such a line of reasoning in some ways resembles the way we think about decision-making about people's health. Bioethics puts a great emphasis on patient autonomy precisely because any medical procedure involves certain risks. It is not morally acceptable for someone else to force a patient into undertaking a medical procedure because allowing the patient to make their own choices and to take on responsibility for the risks involved in undergoing a certain procedure is integral to respecting their autonomy. Just as we ought to respect a patient's autonomy in the medical context, we ought to respect people's intellectual autonomy to form their own opinions.

By writing Mill's Infallibility argument in standard form, we can see how epistemic and moral considerations intermix to support Mill's conclusion:

1. Each person **ought to** form their own opinions and holds responsibility for the risks of being mistaken.
2. We **ought not** impose an opinion onto others unless we have absolute certainty that we are infallible in forming it.
3. No matter our level of confidence, we **cannot** assume infallibility of our own view because we **can never** exclude the possibility that evidence might arise in the future that will prove us wrong.
4. By prohibiting the discussion of someone else's opinion, we assume the infallibility of our own view.
5. We **ought not** and **cannot** prohibit the discussion of someone else's opinion.

Mill's conclusion forbids imposing opinions onto others on both epistemic and moral grounds. While premises 1 and 2 outline moral responsibilities we hold in our own and other

people's belief-formation process, premise 3 arises out of skepticism towards our ability to attain epistemic certainty. Furthermore, this standard formulation of the argument shows that Mill does not require everyone to adopt radical skepticism towards all of their opinions: because we carry responsibility for our own mistakes, we are at liberty to form our own beliefs even without having absolute certainty. When imposing ideas onto others, however, we ought to abide by a higher standard.

b) False Beliefs

Even though Mill's first argument effectively establishes that all opinions ought to be discussed since we cannot be certain they are not true, Mill goes on to strengthen his case by explaining that even if we were absolutely certain of an opinion's falsity, it would still be a mistake to prohibit its discussion. In his defense of the value of false beliefs, he lays out a view of the nature of truth and knowledge. According to Mill, to unwittingly assent to a proposition does not mean knowing the truth: "Truth, thus held, is but one superstition the more, accidentally clinging to the words which enunciate a truth" (103). Rather, knowing a true proposition requires an understanding of both the positive arguments grounding it and the negative arguments that address attempts to falsify it. Moreover, being justified in holding an opinion necessitates a capacity to defend one's own belief. To do so, one must know the arguments of those who think otherwise and have the ability to point out why they are false: "when we turn… to morals, religion, politics, social relations, and the business of life, three-fourths of the arguments for every disputed opinion consist in dispelling the appearances which favor some opinion different from it" (104) Without knowing what the dissenters think about a subject and being able to explain why their arguments fail to hold water, we cannot claim to know and understand our own view in a meaningful sense. Not merely tolerating but engaging with false opinions thus bears insurmountable value for our

own intellectual prowess: “So essential is this discipline to a real understanding of moral and human subjects, that if opponents of all important truths do not exist, it is indispensable to imagine them, and supply them with the strongest arguments which the most skilful devil’s advocate can conjure up” (105). Therefore, even if we are certain a view is false, we ought not to censor it as its falsity helps us better to understand the truth.

c) Partial Truths

To conclusively establish the importance of free discussion, Mill offers a third argument that supports the value of partial truths. He argues that opinions rather than being true or false often fall into the gray area of partial truths: “when the conflicting doctrines, instead of being one true and the other false, share the truth between them; and the nonconforming opinion is needed to supply the remainder of truth, of which the received doctrine embodies only a part” (112). Speaking of degrees of trueness might seem like an illegitimate epistemological move and it is not what Mill is doing. Rather, Mill recognizes that people’s views and opinions on complex issues are made up of a bundle of propositions which rarely contain no falsehoods. Since our faculties of judgment necessarily depend only on the limited number of facts and experience known to us, we are prone to error. To escape biases of subjectivity and attain anything close to objective knowledge, we must thus consult others and through their testimony access data that helps us reach a better judgment. Precisely our epistemic codependence leads Mill to argue that any opinion—especially the non-conformist one—ought to be freely discussed as it may contain a nugget of truth that we would otherwise have overlooked: “every opinion which embodies somewhat of the portion of truth which the common opinion omits, ought to be considered precious, with whatever amount of error and confusion that truth may be blended” (112). If we honestly examine our faculties of reason and consider them alongside the immeasurable totality of what there is to

know, we can appreciate the folly of discarding any fragment of truth merely because it comes wrapped in falsehoods.

III. An Overview

Mill may have risen to such prominence through defending free expression precisely because he skilfully integrates both the instrumental and intrinsic reasons for the value of free expression. He offers a compelling analysis of the fundamental nature of expression that presents expression as more than a mere other-regarding attempt at communication and establishes its role in enabling individuals to think for themselves and have rich internal lives. As such, Mill offers a reinforced and upgraded version of the Lockean view since he emphasizes not merely the intrinsic value of being able to speak freely, but also the worth of being exposed to various opinions of others. Furthermore, Mill's view presents an upgrade to the historical arguments that recognize merely the instrumental worth of free expression. Whereas the Greeks already viewed discussion as a way to improve decision-making, their eagerness to shout down and condemn non-conforming views demonstrates an absence of appreciation for dissenting voices. By defending the didactical value of falsities and partial truths, Mill puts together a case for embracing free expression that transcends historical boundaries as it proves itself as relevant now as it would have been in Mill's, Locke's, or ancient times.

B. Freedom of Expression from an Evolutionary Perspective

Various arguments in favor of freedom of expression in the history of philosophy successfully establish the importance of said liberty for the functioning of our democracies, societies, and minds. However, the importance of free expression is not exhausted by considerations from humanists and social scientists. Considering how evolutionary principles apply to freedom of expression grants us a deeper understanding of the value and the dangers of freeing speech. It leads us to a rather shocking conclusion that the progress of the human species and to an extent our survival depends on freedom of expression. After explaining the theory grounding this section, I will examine how the principle of natural selection and multi-level selection theory apply to ideas and argue that they both suggest that it is in our interest to allow for variability of ideas. I will then address considerations from memetics as a potential reason for censorship. Memetics will lead me to refine my conclusion by proposing that valuing variability of ideas is of much greater evolutionary value to us if paired with deliberate efforts to promote a widespread use of truth-discernment techniques.

In daring to think about freedom of expression from an evolutionary perspective, I am implicitly asserting that evolution has a say in what has traditionally been perceived as a legal or philosophical rather than a scientific topic. My writing leans on the fundamental conjecture of universal Darwinism proposed by Richard Dawkins: natural selection does not merely work on genes, but on any imperfect replicator (Dawkins). In other words, anything that replicates in a way that sometimes by accident or deliberately generates multiple versions of itself evolves overtime because the principle of selection favors its more successful versions. Success of imperfect replicators consists of their ability to spread in their environment, thus maximizing their own survival. Applying the selective principle known in evolutionary theory as natural selection to reproducing things other than genes generates interesting

theories about the development of various aspects of human culture. With the help of universal Darwinism, we can for example explain why ice cream became a sweet rather than a savory dish, why Facebook did better in drawing in users than MySpace, or why Beyonce's music attracts more listeners than Stravinsky. Today, however, I aim to apply evolutionary principles to ideas. Indeed, unlike genes, ideas do not replicate themselves only in passing from one generation to another but rather spread to anyone listening to or reading their expression. They mutate because of our faulty memories and our tendency to knowingly or unknowingly adjust ideas to our worldviews before sharing them further. Because they replicate and mutate, however, they are according to universal Darwinism subject to natural selection.

The most straightforward way of applying natural selection to ideas suggests that just as genes, the ideas selected for are the ones that best contribute to their survival. Because we are the vehicles of an idea's replication, our survival increases the chances of us replicating the idea, thereby often rendering the most replicated ideas the ones that most contribute to our survival as well. The link between the replicatory success of an idea and its contribution to our survival can be easily seen in simple examples, for example, in lessons commonly taught to children such as that you ought to be careful when interacting with strangers. Similar trend lies in the successful spread of ideas about effective wound-treatment after the discovery of the connection between lack of hygiene and infections. Many such ideas considered as common knowledge made their way into our mental inventories because the truths that they communicate such as "stranger danger" and "germs bad" directly contribute to our survival. At first glance, it might seem that ideas selected for are mostly the ones communicating something true about our environment—to best adapt and survive, we must after all know the truth about the conditions putting us in danger. However, while a great deal of ideas aids us in

surviving because they are true, there are many equally helpful, successfully replicating ideas that are quite likely false. For example, most of us hold certain ideas about ourselves and our loved ones as having value, being important, and significant even though we and our loved ones are likely to be at the best average or unremarkable and at the worst downright unsuccessful at most things we do. Ideas about afterlife and divinity occupy a similar function of motivating us to keep living regardless of their truth status. In brief, natural selection is no Galileo dedicated to discovering the truth, the whole truth, and nothing but the truth. On the contrary, the process of natural selection at an individual level discriminates against ideas for their utility rather than their truth.

To establish the importance of increasing the variability of ideas for discovering truth, we must shift the focus of our analysis and examine the way natural selection of ideas operates on a group level. There are plenty of ideas that stick around even though they do not directly contribute to an individual's survival such as the beliefs that we must fulfill our civic duties, ought to feel guilty if we kill a stranger, or that everyone should study arithmetic. To explain the presence of such sentiments we must apply the Multilevel Selection theory that claims that natural selection does not merely select for ideas and genes in as far as they increase the survivability of an individual, but also in as far they increase the survivability of a groups because "groups are most successful (and, thus, able to accomplish their goals and outcompete other groups) when they are able to suppress self-serving behavior that harms the group and to encourage cooperative and altruistic behavior that serves the group" (Seaman & Wilson, 1029). In other words, prosocial behavior and sentiments occur so frequently because belonging to a successful group bolsters an individual's (and his genes') chances of survival and reproduction to a greater extent than endowing the individual with good genes and letting him fend for himself.

Initially, it may seem that prosocial ideas that survive because they contribute to the survival of a group are subject to the same kinds of epistemic concerns as the one contributing to an individual's survival—natural selection again selects for what is useful rather than what is true. However, on a group level much worse long-term evolutionary consequences await groups entertaining false beliefs. For example, a cluster of ideas very popular amongst Shakers, an 18th century American sect, bade people that they all ought to live in celibacy. Unsurprisingly, Shakers and their ideas became virtually extinct. I do not mean to imply that all groups that face evolutionary failure and go extinct do so due to epistemic error, nor that all the victors of history were right. My proposal is much more modest, namely that because extinction on a group level often comes as a result of lacking important information, there is a deep-set evolutionary incentive for groups to attain the knowledge of truth. Healing illnesses, defeating enemies, and ensuring resources for survival all depend on knowledge and without a doubt increase a group's reproductive fitness. Because of such high stakes, groups tend to develop systematic methodologies of storing and propagating knowledge that we refer to as science and education. Steve Stewart-Williams in *The Ape that Understood the Universe* supports such a conclusion by describing the development of science as "an evolutionary process within the realm of ideas," (229) one depending on the Darwinian principles of variation and selection. He outlines the reasons which suggest that natural selection because of the way it functions on a group level enables a spontaneous development a collective truth-producing mechanism:

> "Scientists propose competing theories about the nature of the universe (variation), and then cull those theories that don't match what they see in the world and in the lab (selection)…. In effect, the scientific method establishes a struggle for existence among theories, which results ultimately in the survival of the fittest theories: those that best

explain the facts. The end result is that our theories evolve – step by slow step – toward greater and greater accuracy" (229)

If on a group level natural selection selects for ideas that promote the survivability of the group and the survivability of the group is on the most part dependent on the group's ability to acquire true beliefs, then letting selection run its course ought to gradually bring us to the truth. The best way to contribute to our group's evolutionary success is to aid it in its quest for truth by freely and abundantly generating a wide variety of ideas so that the best ones can be selected for. Freedom of expression ensures variability which gives natural selection more to choose from thus ensuring that the best ideas come to light. Huzzah for freedom of expression!

Before I pull out the champagne and toast to the health and longevity of ideas in support of freedom of expression, however, I must face a serious complication without which my argument has little practical application. Namely, there are ideas that are so good at spreading themselves within a group that they are favored by natural selection simply because of their circulatory success. I will follow the terminology Dawkins introduces in *The Selfish Gene* (1976) and call ideas or units of culture selected for their success in "selfishly" propagating themselves memes. In words of Stewart-Williams, memetics forwards the idea "that, like genes, memes are subject to natural selection, and that selection favors "selfish" memes—memes that, through accident or design, are good at getting themselves replicated and keeping themselves in circulation in the culture" (222). If ideas most widespread and alive at any given time may have been selected not for how much they help us in survival, neither for how true they are, but for how infectiously and rapidly they spread through an existing population that presents a problem for the freedom of expression argument–the most fashionable rather than the truest idea wins. Selfish memes are the equivalent of crocs clogs:

crocs were extremely popular within certain groups in the past decade. Their popularity, however, cannot be attributed to them bringing a revolutionary advantage into the shoe industry—breathable and durable shoes have existed before and after. Moreover, they are in no way aesthetically pleasing (beauty in this analogy serves as the truth of the fashion industry). They were a fleeting fashion trend that enjoyed its five minutes of fame because of good marketing and a certain quality that in a correct historic moment succeeded in infecting the minds of consumers with the idea that they must buy a pair. Like Crocs clogs, some selfish memes with little epistemic or survival value become popular merely due to being infectious enough at a certain historic moment. They range from harmless but deeply erroneous value judgements such as "buying my friend a pair of crocs is a great idea," to more serious misconceptions such as "all women are bad at mathematics, chess, and leadership," and lastly include selfish memes with horrific consequences such as the super-spreading ideas that paved the way for the Holocaust in the 20th century. Threatened by such infectious and potentially harmful memes, we cannot rely simply on increasing the variability of ideas and letting natural selection bring us to truth.

Indeed, we cannot. While some may see selfish memes as a reason for censorship, however, I argue that controlling memetic evolution with educational techniques is by far the better way. For as long as we have lived and competed as groups, we have developed techniques for truth-discernment that help contain and prevent spreads of selfish memes. Stewart-Williams refers to these as "cultural mechanisms that reliably favor truth over catchiness" and lists "critical thinking, careful observation, peer review, open discussion, independent replication, and the rejection of authority, tradition, and revelation as reliable sources of knowledge" (268). The scientific method as the culmination of such techniques therefore helps us curb the reign of selfish memes and stick to the truth. The promotion of the

scientific method through education best helps us to stay on the evolutionary track of survival and truth and prevents selfish memes from slowing us down or drawing us into extinction. Fighting selfish memes with education is preferable to censoring them because censorship leaves their infectious nature and spreading potential intact. Education, on the other hand, by revealing the vacuous selfishness of a selfish meme damages its capacity to spread. By reminding us we have a vested interest in pursuing truth, the scientific method gives us the motivation and the skills necessary for controlling the evolution of ideas in our society.

We need freedom of expression to give our pool of ideas sufficient variability so that the most useful and the truest ideas can come to light. However, since every vessel of ideas has the potential to succumb to harmful yet superspreading kinds of selfish memes, we need to pair freedom of expression with vested efforts to help thinkers acquire the skills necessary for applying the scientific method to separating truth from falsity. We may even be evolutionarily inclined to develop reliable truth-discernment techniques overtime whether or not we consciously pursue it: recent analyses of historical examples (Henrich, 2022) and experimental studies (Thompson et al., 2022) support the notion that evolutionary principles of variation and selection play important roles in developing, filtering, and preserving complex human knowledge systems. Whether it develops spontaneously or not, careful, analytical, and evidence-based reasoning is necessary for the success of our species as it helps us cripple the spreading mechanisms of infectious but false ideas while increasing the evolutionary success of the truth. Censorship merely shuts our eyes to the threat of infectious and false ideas and drives them underground where they await favorable historic circumstances in which they will once again wreak havoc on humankind.

PART 2: WHY EXPRESSION IS DIFFICULT TO REGULATE

A. A Brief Overview of Legal Restrictions of Expression

1. International Law

On an international level, freedom of expression and its limitations are outlined in Articles 19 and 20 of the International Covenant on Civil and Political Rights (ICCPR). According to Article 19 of ICCPR, freedom of expression "carries with it special duties and responsibilities" and thus may be restricted but only by restrictions that "are provided by law and are necessary," namely "for respect of the rights or reputations of others" and "for the protection of national security or of public order, or of public health or morals" (TWG, 4). Even though some may argue that Article 19 in allowing restrictions of expression to protect public morals might already be too expansive, Article 20 goes even further as it requires the prohibition of "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" (TWG, 7). Among the ICCPR's signatories, US, Belgium, and Denmark have made reservations with respect to Article 20, US on the grounds of it being unconstitutional. The ICCPR has the United Nations Human Rights Committee to oversee its implementation. As a report of the Transatlantic Working Group points out, however, the US in particular "tends to be relatively non-receptive to the influence of international law" (TWG, 4).

2. The United States

The US is unreceptive to international law when it comes to freedom of expression because in no other country expression is as protected as it is in the US with the First Amendment. The First Amendment protects even expression deemed offensive, immoral, discriminatory, hateful, etc. In the first half of the 20th century, the Supreme Court could constitutionally punish speech that passed a bad tendency test based on English common law.

It allowed the Court to punish expression on the vague notion that "the natural and probable tendency and effect of the words are such as are calculated to produce the result condemned by the statute" (Gibson). In other words, if expression has even a tendency to incite or cause illegal activity it may be stifled. In practice, the bad tendency test was used to eliminate political dissent. In mid-20th century, the bad tendency test was thus replaced with the clear and present danger test determining that it is only permissible to stifle instances of expression that in the words of Justice Oliver Wendell Holmes presents "a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent" (Parker). Furthermore, laws restricting speech such as for example libel and slander laws must abide by the viewpoint neutrality principle that prohibits laws from discriminating against speech based on it containing a subject matter that governmental officials disfavor.

3. The European Union

In the EU, the laws restricting expression vary from country to country. Nevertheless, Article 10 of the European Convention of Human Rights (ECHR) protects free expression and like ICCPR emphasizes that the exercise of free expression "carries with it duties and responsibilities" and as such may be restricted "in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary" (TWG, 5). ECHR provides a lot of grounds for limiting expression with a lot of vague diction and the trend improves little in the regulation of individual member states.

A 2017 report of ARTICLE 19, for example, looked at the legal framework and practices related to 'hate speech' regulation in Austria, Germany, Hungary, Italy, Poland and the UK. They have found "widespread deficiencies" in national regulatory frameworks from

issues with their "compatibility with applicable international freedom of expression standards" to "inconsistencies in the application of existing legislation" (4). The flaws in legislation, according to the report writers, have serious consequences as they "render the legal framework open to political abuse, including against precisely those minority groups that the law should protect" (4). While in the US, some voice complaints and frustration at the First Amendment for too narrowly construing conditions of expression restriction, the EU's frustration has largely to do either with laws restricting too much speech or restricting speech in an unfair, politically-motivated way.

B. What Speech Regulation Should Look Like

The biggest difference between the US and EU speech regulation are the grounds upon which regulation is permitted. Even the most stringent defendants of freedom of speech agree that expression ought to be restricted when its restriction prevents a directly caused and significantly pernicious harm—such a standard is reflected in the clear and present danger test. While there is disagreement on what kind of expression qualifies for restriction on such grounds, it tends to focus on the prevention of serious harm. Many, however, believe that not only harmful but offensive and hateful speech warrant restriction—the most illustrative example of the latter are hate speech laws in Europe. Setting aside the question of what kind of legal practice generates most desirable results, there is a difficult theoretical discussion to have about what kind of expression warrants regulation and why. A comprehensive examination of all the factors relevant to defining the ideal limits of free expression would require a much lengthier analysis that would include topics such as the nature of language and causality, the nature and role of the government, and would posit a definitive definition of terms such as harm, offense, and hate. I will likely not be able to do justice to all the nuances of the matter. I hope, however, to offer a compelling argument for why principles regulating speech should be construed as narrowly as possible in the way more akin to Mill's Harm Principle than modern, especially European, expansive speech-regulating legislation.

I plan to examine harm, offense, and hatred as grounds for restricting expression and interrogate the main challenges that plague those who wish to define what kind of expression qualifies as harm-causing, offense-inducing, or hate-spreading in a way that warrants restriction. In discussing harm, I will build on Mill's Harm Principle to extract three standards that I will defend as both necessary and sufficient conditions for regulating speech on the grounds of either harm, offense, or hate: A. The Severity Standard, B. The Direct Causality

Standard, and C. The Last Resort Standard. After outlining what kind of speech satisfies the three standards when it comes to harm, I will turn to offense and interrogate it through the same principles. I will first draw on Feinberg's position in *Offense to Others* to discuss the way the three standards are applicable to offense and then address an important objection that regulating offensive expression is a matter of setting community standards and does not need to fulfill such strict standards that may more sensibly apply to harmful speech. In discussing hate, I will draw inspiration from Jeremy Waldron's argument from *The Harm in Hate Speech* to make a case for justifying the regulation of expression even when it does not satisfy the Severity, Direct Causality, and Last Resort standards. I hope to demonstrate that relative to the importance of freedom of expression, only the expression satisfying the three standards warrants regulation. Any attempt to disregard or loosen up a standard becomes overinclusive and opens up the door for misapplication that leads to severe encroachments on fundamental human liberties.

1. Harm

Even the most passionate advocates of free speech recognize the importance of restricting expression that leads to harm. At the beginning of *On Liberty*, for example, Mill identifies harm to others as the only valid reason for governmental interference with individual liberties: "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others" (80). Known as the Harm Principle, this guideline comes as a result of Mill's arguments for the value of free speech and highlights the importance of free expression—no matter how offensive, upsetting, mean, and disgraceful expression can be, it may not be infringed upon unless it causes harm to others. To apply the Harm Principle to real life situations, however, it requires further elaboration. Most importantly, its vague initial formulation calls for defining harm and

for discussing what kinds of harms warrant restriction. If defined too loosely, the Harm Principle could be seen as suggesting that restriction of speech is due even if it causes only the slightest emotional discomfort i.e when a mother tells her daughter that she is behaving poorly and by doing so "hurts her feelings." If restricted too narrowly, however, it may only allow for prohibiting speech that causes an exorbitant amount of physical harm such as for example an exclamation that incites a mob to rampage. While freedom of speech is a very important liberty, it would be too inconsiderate to deny protection to victims of harsh verbal abuse or to those who suffer a less tangible harm such as serious harm to reputation inflicted by slander or libel. To strike the right balance, I will begin by discussing the definition of harm and then forward three standards for determining whether a harm caused by expression is of the kind that warrants restriction.

I. What Counts as Harm?

Today we recognize several different kinds of harm. The main most clearly distinct two types are physical harm such as, for example, bodily injury and psychological harm that manifests itself in a decrease in mental health. However, there are also more abstract harms such as the previously mentioned harm to one's reputation, honor, credibility, and so forth. While the harm-types differ greatly, they all have in common the fact that the victim of harm suffers a setback to his interest. Therefore, a necessary (though perhaps not sufficient) condition for harm to occur, is a setback to the victim's interests. For example, if someone stabs me, I am harmed because before the incident I was better off with regards to my interest in bodily health. Similarly, when I undergo psychological harm, I come out of an experience worse off with regards to my interest in being of sound and stable mind.

The inclusion of psychological harm under harm has not always been thought self-evident. Mill, for example, in explaining his Harm Principle focuses mostly on the

occurrence of concrete, clearly recognisable physical harm. In the third chapter of *On Liberty*, Mill points out that in some circumstances, speech can no longer be protected as a mere self-regarding action: "even opinions lose their immunity, when the circumstances in which they are expressed are such as to constitute their expression a positive instigation to some mischievous act" (121). In such situations opinions become actions because it is evident from the circumstances of their expression that they do not aim to spur discussion but rather cause an action. As an example, Mill offers an expression of the opinion that "corn-dealers are starvers of the poor" (121) expressed to an excited mob in front of a corn-dealer's house. The same opinion may be printed in newspapers or brought up in a debate, however, due to the specific circumstance of its expression, it becomes as significant of a causal factor as an action would be. Because it causes an outburst of violence, expressing such an opinion violates the obligation to do no harm and may therefore be prevented or punished. In contrast, Mill is not eager to apply the Harm Principle to psychological distress caused by critique and disagreement: "if the test be offense to those whose opinion is attacked, I think experience testifies that this offense is given whenever the attack is telling and powerful" (118). Admittedly, it is unclear whether Mill does not see the Harm Principle as applicable to offense because he does not recognize psychological harm or because he merely believes that most often non-physical harm is not of sufficient severity to warrant speech-restriction. On either reading, however, the Harm Principle presents harm as a necessary but not sufficient reason for limiting speech.

II. What Kinds of Harm Count?

When a harm occurs as a consequence of speech, it does not mean that said speech may be restricted. To define the conditions that expression must satisfy in order to warrant regulation, I will lay out three standards any instance of expression must satisfy in order to

warrant prohibition. They are inspired by the essential traits of Mill's corn-dealer example that render the expression "corn-dealers are starvers of the poor," (121) a clear example of speech that may be interfered with to prevent harm to others.

A. The Severity Standard

In the corn-dealers example, a lot is at stake if the inciteful speech is permitted. An enraged mob will not only cause significant material damage but will likely murder the corn dealer and his family. The harm that the state would be preventing by prohibiting expression of a speaker inciting a mob is significant and severe. As Mill's case suggests, a Severity Standard of some sort is necessary to rule out trivial instances in which the harm afflicted is not proportionate to the significance of infringing upon someone's liberty to express himself. For example, voicing a critique of Kim Jong-Un's way of governing North Korea may indeed tarnish the dictator's reputation, but such a harm is in no way severe enough to warrant a country-wide ban on expressing political opinions. Freedom to express oneself is too important of a liberty for an individual's personal autonomy and a healthy public discourse to restrict it when it causes only a minor harm. Indeed, it is a matter of interpretation what passes as severe harm. To the rulers of the North Korean regime questioning the dictator may very well seem severe. In order to mitigate such relativity of perception, it may be best to define as severe only those instances of harm in which the setback to the victim's interests is such that it causes a significant disruption to the harmed party's regular course of life.

B. Direct Causality Standard

Another important aspect of the corn-dealer's example is that due to the temporal and spatial proximity between the objectionable expression and the occurrence of harm there is no doubt that the mob's violence was directly caused by inciteful speech. Direct Causality Standard requires harm to be directly caused by the speech in question—most often the

standard is satisfied when harm occurs immediately after the speech and in a close spatial proximity to the speaker. Such a standard helps rule out cases in which the causal connection between an expression of an idea and an occurrence of a harm is at best questionable. If, for example, the speaker of the inciteful sentence was to express it in a passionate rant to a fellow villager over a pint of beer, and then said villager was to go and murder the corn-dealer later that same evening, we would have little ground to restrict that same speech—the most proximal cause of the harm is the villager's decision to murder rather than the speech that inspired him to do so. Historically, the government and religious authorities have frequently prohibited expression of certain ideas which they believed might be causally connected to a certain future harm: for example, Goethe's novel *The Sorrows of Young Werther* was for a time taken out of circulation because it was believed to cause a spike in suicide in young people. As recently as in the first half of the 20th century, the US Supreme Court could constitutionally punish speech on the vague notion that it may cause some harm or in other words, that "the natural and probable tendency and effect of the words are such as are calculated to produce the result condemned by the statute" (Gibson). In practice, this loosened standard became an excuse to prosecute political opponents with anti-war sentiments. The Direct Causality Standard is therefore necessary not only because without it, the causal connection between expression and harm becomes too weak, but also because removing it allows for political abuse.

C. The Last Resort Standard

Finally, the corn-dealer example offers a case in which prohibiting or preventing expression is clearly the last method available for preventing the occurrence of harm. For example, the excited mob in Mill's example is in a very emotional state—enraged and excited it waits only for someone to yell "Go!" and give them an excuse to attain a premeditated goal

and punish the corn-dealer. Their rampage does not come as a consequence of a calm and rational decision to harm the corn-dealer, *because* a speaker brought to their attention that the corn-dealer is harming them. Because the deliberative content of inciteful speech does not matter to the mob, no amount of debate would be able to mitigate the effects of the exclamation "corn-dealers are starvers of the poor" (Mill, 121). Since prohibiting the expression of said idea is the only way of preventing the mob from rampaging, prohibition is permissible. The Last Resort Standard dictates that only when other options such as debate would be ineffective, expression may be curtailed by censorship. Without a respect for such a standard, regulation of speech rather than discussion becomes a go-to method of generating compliance. As censorship prevalent on social media platforms such as Facebook and Twitter may suggest, the prohibition of speech appears to be an easy way to deal with potentially harm-inspiring opinions that enter the public forum. It is by no means, however, better than exposing opinions that may inspire violence as false by providing better, more persuasive arguments against them. Only by treating censorship and prohibition as a last-resort tool, we honor the value of freedom of expression and avoid stifling it in pursuit of an unachievable level of safety.

III. Objection: We Should Err on the Side of Safety

When it comes to discussing harms caused by expression, many might have an intuition that it is better to be safe than sorry and overregulate rather than underregulate and allow harm to occur as a result of expression—even if expression is not the most direct cause of not even necessarily all that severe a harm. In other words, some believe that we should prohibit and censor a bit more expression than strictly necessary to prevent harm. The idea is that making someone shut up about certain kinds of topics that are likely to lead to violence is a small price to pay to avoid the suffering that might ensue. Moreover, very often the

potentially harm-causing expression has little informative or enlightening value for public discourse: why should people be free to express, for example, a disdain over existence so profound that it inspires suicide in their readers or, for instance, to rage against authorities or certain groups of people with a language so powerful that it incites their listeners to spread hatred and violence in their own lives? Surely, not much value is lost in sacrificing certain topics, ideas, or words for the common good?

While the desire to ensure greater public safety is admirable and natural, sacrificing freedom to discuss some ideas or topics in pursuit of public safety is ill-conceived. It violates individual autonomy in much the same way as if the government was to constantly monitor all its citizens' conversations so as to prevent any scheming of illegal activity. As autonomous human beings with rights and dignity, we ought to be free to think and discuss even offensive, daring, morally comprehensible, and "dangerous" ideas. No one should have the right to interfere with the most fundamental processes of our individuality such as thinking and expressing our thoughts. Because of its importance for the exercise of individual liberty, we must not regulate expression based on the topic and content but only based on the situational aspects of expression that render it too great a threat to public safety to tolerate. As Mill points out in *On Liberty*, "no one pretends that actions should be as free as opinions" (121). In censoring expression, we thus ought to stick narrowly to the kind of expression that because of its situational context attains the nature of an action as its direct consequence is an occurrence of severe harm that nothing but the expression could cause because nothing, but the censorship of expression can prevent it.

IV. Limitations of the Three Standards

I propose the Severity, Direct Causality, and Last Resort standard as the best compromise between addressing the unwanted consequences speech may have while keeping

the most expression possible out of the reach of governmental authorities. Nevertheless, they cannot escape the two most problematic aspects of any law or standard that seems to plague most regulations pertaining to speech censorship—the issue of being so vague that they can be applied in a way that makes them overly-restrictive and the issue of not being restrictive enough.

One way to counterbalance vagueness is by examining why certain examples satisfy or do not satisfy the standards and therefore clarify the ideal ways of interpreting what is "severe," "a direct cause," and "a last resort." Comparing, say, Mill's corn-dealer case with the prohibition of regime-skeptical political expression in North Korea, however, only clarifies the more extreme ends of a spectrum: speech that clearly may be restricted and speech that obviously ought not to be. In between the two, however, remains a gray area of more everyday cases in which it is admittedly extremely difficult to figure out whether the speech is a direct cause of a severe enough consequence that cannot be prevented in other ways than censorship. The vagueness that prohibits us from clearly determining at what point the speech brings about a severe and proximal enough consequence is undoubtedly a drawback of my standards, but it also makes them flexible enough to be adjusted to a wide variety of expression-related situations. The imperfections of the judiciary system and fast-changing culture norms indeed render any vagueness a potential loophole to abuse. Nevertheless, I believe that the best way to safeguard freedom of expression for the future generation while serving justice when a speech-related injury occurs, is to make sure that we can answer in the affirmative to these three questions: 1) Is expression truly a direct cause of the consequence? 2) Is the consequence of expression severe? and 3) Is censorship our only option?

On another note, I realize that there are certain situations in which speech may be broadly viewed as reprehensible enough to be censored or punished but does not satisfy my three standards either because the causal link is too weak or because the consequence it brings about is not clearly or necessarily severe. For example, after school shootings we often hear of unregulated speech platforms such as 4chan and 8chan on which shooters published their manifestos and met others with similarly dark thoughts whose encouragement might have influenced them to commit an atrocious act. It seems very appealing to propose censorship and regulation of those platforms even though it is not clear that the platform culture and users were a direct cause of the individual's behavior. The severity of the crime is so high that it seems very tempting to disregard the Direct Causality standard and pursue regulation. A similar problem arises when we consider un-targeted expressions of racial slurs and highly discriminatory discourse such as crude cartoons or scribbles on bathroom stalls and other public areas. Such forms of expression are generally of extremely low discursive value and wickedly exploit a painful history of racism to propagate discriminatory sentiments. However, while deeply offensive, they are not necessarily always demonstrably causing a severe consequence—often that apparent lack of a consequence comes due to black people's resilience in face of prejudice. It seems, however, that even if the targets of such speech have learned to dismiss it and the Severity Standard is left unsatisfied, racial slurs are so morally reprehensible that we need not tolerate them regardless of people's sensitivities to them.

Such examples demonstrate the primary difficulty of defending narrow speech-regulation standards—there are going to be situations you leave out that you might prefer not to. Many would rather see more allowing standards even if that means there are cases of over-regulation and overly eager censorship. However, the historical arguments for

freedom of expression that I tried to convey in the first part of my thesis by discussing Ancient Athenians, Locke, and Mill in my opinion present a forceful and compelling case for the instrumental and intrinsic importance of freedom of expression. Before one convinces me that over-regulation is preferable to under-regulation, I will need to see not merely contemporary examples but a persuasive refutation of the historical foundation on which my case stands.

2. Offense

Even though abandoning the three standards to prevent harm may be unwise, there are perhaps other grounds for prohibiting expression that may permit us to sufficiently expand the array of expression that may be regulated. The proponents of expanding the government's regulatory privileges thus argue that not only expression that causes severe harm to others, but also offensive expression that evokes undesirable mental states and unpleasant emotions may be punishable by law. After all, we permit the government to legally restrict our freedoms not only to prevent harm, but also to enforce standards in our community that make our society more pleasant to live in. For example, there are laws regulating public urination and defecation as well as public indecency. If such unpleasant behaviors may be prohibited without much fuss, why would offensive expression be any different?

A wide range of expressions may cause us to take offense: from a person in the cinema loudly making crass comments on the appearance of the lead actress, a teenager blasting rap music with highly inappropriate lyrics on the bus, all the way to someone openly mocking an aspect of our identity e.g. our religion, nationality, or gender. Undoubtedly, offensive expression can be highly unpleasant; it makes us feel negative feelings such as annoyance, disgust, shame, anxiety, or humiliation. In *Offense to Others* Feinberg offers an

argument in favor of regulating offense defined as "the whole miscellany of universally disliked mental states" (1). He forwards the offense principle which states that "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" (xiii). Yet just like with harm, Feinberg needs to add a series of tests that offensive expression needs to pass in order to warrant restriction. After presenting the difficulties with differentiating offense from harm and laying out Feinberg's Offense Principle, I will show how the conditions set by Feinberg correspond to the Severity, Direct Causality, and Last Resort Standard that I set out as guidance in restricting harm.

I. The Relationship Between Harm and Offense

Some argue that we should regard offense as a genre of harm and need not consider it separate from it. Others like Feinberg, however, wish to strictly distinguish between the two. Feinberg wants to say that no offense should be viewed or understood as harm: "Continued extreme offense…can cause harm to a person who becomes emotionally upset over the offense, to the neglect of his real interests. But the offended mental state in itself is not a condition of harm" (3). He sees his principle as one demonstrating that certain offensive, but not harmful experiences may nevertheless be restricted: "we can rightly demand legal protection from them even at the cost of other persons' liberties" (10). While a strict separation between harm and offense may be useful for providing a legal principle, the two categories are closely related in reality: harm often accompanies offensive situations even though the causal connection between the offensive expression or behavior and harm may often be less clear.

To dig deeper into the relationship between harm and offense, let us consider one of the most horrifying imaginary scenarios intended to test our intuitions that Feinberg offers in

his first chapter. Feinberg urges us to imagine an event taking place on a crowded bus that we must take to get to our destination which means that we cannot escape the situation by switching seats or leaving: “A passenger with a dog takes an aisle seat at your side. He or she keeps the dog calm at first by petting it in a familiar and normal way, but then petting gives way to hugging, and gradually goes beyond the merely affectionate to the unmistakably erotic, culminating finally with oral contact with the canine genitals” (12). If I put myself in such a situation, I can vividly imagine experiencing intense emotions of revulsion, horror, and indignation. But despite the intensity of these negative feelings for the duration of the incident, the fact that I was in an offended state of mind is not sufficient to establish that I was harmed.

As mentioned earlier, a necessary (though perhaps not sufficient) condition for someone to harm me, is that he sets back my interests so that with regard to a specific interest, I am worse off after the harm occurs than I was before it occurred. In the case of extreme offense, the type of harm on the table is psychological harm which would require me to come out of an experience worse off with regards to my mental health. Importantly, mental health does not pertain to feeling a certain way in any given moment, but rather to having an inventory of mental skills and emotional mechanisms that help me face the challenges of my life with responses proportionate to the challenges. For example, both a mentally healthy and a mentally unhealthy person may panic when on a plane about to crash, but only someone whose mental health has been harmed will panic when a stranger in a parking lot loudly bangs the car door shut. Merely going from happy to unhappy is not sufficient for establishing psychological harm; feelings are transient and superficial whereas the health of our psyche runs deeper and changes less quickly. Therefore, if instances of offense are to be considered as the cause of psychological harm, it is not because they provoke negative

feelings, but because they provoke negative feelings of such nature and intensity that they cause a decrease in a person's mental health.

II. The Offense Principle

While Feinberg does not see an overlap between offense and harm, he constructs his principle for offense-regulation so narrowly that the pool of offense-evoking experiences we are left with consists almost exclusively of instances in which offense coincides with harm. According to Feinberg, there are several factors in need of consideration before a given instance of offense warrants punishment and prohibition such as the objective magnitude of the offense, whether it is possible for the offended party to avoid offense, whether the offended party wilfully sought out the offense-inducing situation, and whether the offended party has an abnormal susceptibility for offense. After explaining Feinberg's four limiting conditions I will discuss their interaction with the Severity, Direct Causality, and Last Resort Standard.

i. Objective Magnitude

In order to satisfy the requirement for sufficient magnitude and objectivity, the offense must be serious enough, namely it needs to be "caused by the wrongful (right violating) conduct of others" (1-2). By setting a magnitude Feinberg hopes to exclude situations of trivial offense that may be emotionally intense but do not violate any rights such as for example the offense taken by an Italian chef when someone suggests eating pasta with ketchup. In other words, Feinberg restricts his principle to instances in which "offense" stands for not a subjective description of a mental state, but an objective condition that results in a wrongful action of someone else: "In the strict and narrow sense, I am offended (or "take offense") when (a) I suffer a disliked state, and (b) I attribute that state to the wrongful conduct of another, and (c) I resent the other for his role in causing me to be in the state" (2).

Similar to how suffering offense that does not occur as a result of a wrongful action ought not be regulated, the offense principle is not applicable to situations in which the person causing the wrongful action cannot be blamed for it (e.g. if a child pees in my pool, I cannot blame him for it even though his actions will put me in an unpleasant mental state and will destroy my private property).

ii. Reasonableness of Avoidability

The offense principle likewise does not apply to situations in which "the reasonable avoidability standard" (26) is not met. Many situations that have the potential to offend us can be easily avoided without us incurring any significant costs. For example, we can ignore a group of demonstrators rather than engage in an argument with them, avoid visiting pornographic sites, and not go to comedy shows of comedians whose humour might offend us. Therefore, we ought not punish offense when the victim could have avoided the incident without incurring great costs: "no one has a right to protection from the state against offensive experiences if he can easily and effectively avoid them without unreasonable effort or inconvenience" (32). Applying this condition importantly rules out the suppression of art, books, or social media profiles on grounds of evoking offense.

iii. The Volenti Maxim

The volenti maxim establishes that someone willingly placing themselves in an offense-evoking situation is not wronged—*volenti non fit injuria*.[3] It excludes the malicious versions of the situations ruled out by the reasonableness of avoidability standard in which the offensive experience is actively sought out by "the victim." The offended party's emotions may be incredibly intense and produced by a wrongful act, but if they were voluntarily sought out no wrong happened: "The offended states induced by such voluntarily

[3] Direct translation: to a willing person, an injury is not done.

undertaken experiences are perfectly real, just as the broken bones incurred by the stunt motorcyclist are perfectly real harms, but in neither case can the victim complain of a grievance" (33).

iv. Discounting of Abnormal Susceptibility

In regulating offense, another factor that needs to be accounted for is the subjectivity of susceptibility to offense. There are people, for example, who will suffer offense at the very thought of certain kinds of ideas being in circulation or behaviors taking place and who are thus in virtue of their predispositions unable to avoid offense. Imagine, for example, a person offended by the mere existence of a homosexual couple living next door, or a student appalled by the thought that a speaker invited to her campus will be sharing opinions she finds immoral. According to Feinberg such cases do not warrant legal protection since "the more fragile our sensitive sufferer's psyche, the less protection he can expect from the criminal law" (34), To illustrate the issue further, Feinberg compares an overly sensitive person to a skittish horse. We would not punish someone whose harmless activity startles such a horse but would rather expect owners of such horses "to keep them away from "startling" activities and to take steps to cure them of their skittishness" (34). In other words, we cannot adjust our society to suit the comfort of every single individual—people with deviating sensibilities have the responsibility to cope with them or adjust them on their own.

III. The Purpose of Feinberg's Conditions

In narrowing down the conditions for regulating offense, Feinberg attempts to take offense-regulation as far as possible from it being a matter of measuring the subjective intensity of an individual's emotion. He only wishes to punish the conduct that could not have been avoided and violates not merely our norms of behavior but our rights in a way that would put most individuals at significant unease. The conditions which he believes ought to

be fulfilled before offensive expression warrants prohibition closely resemble the Severity, Direct Causality, and the Last Resort Standard. First, the Objective Magnitude condition attempts to define the graveness of the offense in a standardized way so as to avoid regulating expression when offense is caused by subjectively upsetting but not objectively wrongful behavior. Second and third, the two conditions mediating the behavior of the victim—Reasonableness of Avoidability and Volenti Maxim—attempt to weed out cases in which the causal connection between the expression and offense inflicted is very weak: both when the victim could have easily avoided offense and when she seeks it out, the offense is self-caused rather than the responsibility of the offender. Thus, they correspond to the Direct Causality standard in so far as they wish to make sure that the responsibility for the negative consequence—be it harm or offense—indeed lies with the speaker. Last, discounting abnormal susceptibilities has much the same result as the Last Resort Standard which dictates that censorship is permissible only when other viable options of lessening the negative result of expression are exhausted. If a person gets offended too easily and could have avoided intense offense by practicing mental resilience and introspection—to become less skittish, if we borrow from Feinberg's earlier analogy—then that is an option that needs to be considered before censorship. With physical harm it would be pointless to demand of people to toughen up their skin or wear protective armor all the time, however, the level of emotional upset that occurs when someone experiences offense is much more malleable and greatly depends on factors subject to change and interrogation such as personal resilience and cultural norms.

IV. Objections

Once we apply the Severity, Direct Causality, and Last Resort Standard that may when it comes to offense be reasonably represented by Feinberg's four conditions, we rule

out as unfit for regulation most of the usual instances of offense that people encounter on a daily basis. Only the most outrageous and repeatedly occurring instances of offensive expression pass such strict requirements: we may, for example, have grounds for penalizing a group of passionate atheists that comes to stand in front of an Islamic Mosque every week to yell at the service-goers extraordinarily imaginative obscenities about the prophet Mohammed with a vehemence that would make any reasonable person uncomfortable. Many less extreme and more commonly occurring situations, such as for example the publication of rude cartoons about the Islamic prophet or Catholic religious figures in magazines and newspapers or people wearing t-shirts with obscene messages clearly aimed to demean women, however, would have to be tolerated as in both of those situations, offense can be easily avoided. And so would tamer instances of offense-inducing expression such as people misusing other people's pronouns or wearing MAGA hats or communist insignias; in such instances, offense can be discounted due to an abnormal susceptibility to hotly debated political topics such as the issue of gender and Donald Trump. As the responses of people on Twitter and other social media suggests, however, many desire to prohibit and punish even not so extreme cases of offense seems relatively prevalent. I will present two arguments for loosening the conditions for restricting offensive expression: an argument for the right of a community to determine its own standards and an argument for the importance of kindness and compassion.

i. Community Standards

Many might disagree with the position that only instances of severely offensive expression that wrongs the offended party on an objectively determinable scale of magnitude qualify for restriction. The subjectivity of offense that motivates setting such a condition, may to some seem the very reason for why we ought not try to set objective measures of

offense, but rather let communities decide themselves what kinds of expression they find offensive and do not want to tolerate. Prohibiting offense, after all, is a matter of making a community more pleasant to live by protecting its members from experiencing unwelcome emotions evoked by offensive expression. Such a goal can be achieved only if the standards for what is offensive are adjusted to fit the subjective experiences of community-members. Whereas harm may be objectively quantified and regulated in the same way across communities, offense ought to be adjusted by each group individually based on community-input. Publications ought to listen to their readers, municipalities to their residents, and countries to their citizens in determining and readjusting the rules about what kind of expression is too offensive to pass. If standards of offense are not determined somewhat democratically and adjusted to people's opinions, there is not much point in having them at all. After all, right-violating conduct is regulated with laws already—the only added value of regulating offense comes if we can restrict the kinds of behavior and expression that is not really wrongful but merely perceived as very unpleasant by a great number of community-members.

While the idea of having a community set its own standards of offense so as to maximize the comfort of its members is appealing, it has several flaws that make it impossible to realize without severe violations of liberty. First, it underestimates the dangers of mistaking a minority of loud and easily offended members for the majority opinion. Regulating offensive expression in a way that accommodates those most susceptible to upset often restricts free speech without a significant benefit to the common good. By giving the most easily upset people validation, such a way of restricting expression may, as a matter of fact, may even increase other people's general susceptibility to offense without any valid reason. Even if the majority in a given community is indeed offended by a specific type of

non-wrongful expression, their desire to feel comfortable ought to not override the right of others to express freely. Take, for example, an extremely liberal suburban community in California in which a vast majority of people would feel horrified and upset if faced with the views of a Trump supporter. Banning Trump-supporting discourse would certainly allow those community members to lead more comfortable lives—at what cost though? An increase of comfort and satisfaction of people ought not override someone's fundamental right to discuss their political opinions and values.

If we let communities pick and choose offensive expression regulations according to what they are comfortable with, we not only unfairly infringe upon individual rights of dissenters, but also potentially sacrifice important lessons about coexisting with different-minded people. Just like an extremely conservative Catholic community might benefit from encountering an Atheist, Californian liberals might benefit from an occasional encounter of a Trump-supporter—regardless of whether the encounter makes the group question or reinforce their views. Non-wrongful offensive expression rarely harms the offended party, but it always exposes her to ideas she disagrees with. Being exposed to ideas that go against one's worldview, while uncomfortable in the short term, is likely to benefit one in the long term. The regulation of offensive expression ought to be as narrow as possible and as objective as possible to counter our tendency to engulf ourselves into thought bubbles.

Giving the majority the power to infringe upon an individual's right to free expression presents a risk incomparably greater to its benefits. Throughout history, expression has been censored under the guise of causing offense. For example, in 1885 a journalist named W. T. Stead exposed a child slave trafficking ring in an article. Whereas those participating in child prostitution were not sanctioned, Stead was condemned to a year-long imprisonment for "writing about an indecent subject" (Feinberg, 5). Communities can inevitably enforce

standards of expression to an extent by applying social pressure on on-conformers, but they ought not be allowed to infringe upon their members' fundamental liberties any more concretely.

ii. The Value of Kindness

An alternative objection to the very strictly limited regulation of offensive expression that Feinberg and I advocate for may counter the idea that we ought to discount people with abnormal susceptibilities to offense in determining what expression is offensive enough to deserve prohibition. One may argue that defining what counts as abnormal is in itself extremely subjective and dubious. For example, if we view as abnormally susceptible to offense only those individuals who deviate from the median level of upset in a given community at a certain time, who is to say that their upset is not more adequate than that of everyone else? In Nazi Germany, for example, the median level of offense at antisemitic expression was likely quite low and yet we might want to say that those "abnormally susceptible to offense" had better instincts in said case. Alternatively, if we define abnormal susceptibility to offense as a deviation from a less specifically determined median, are we to include the upset levels of all communities from all cultures and eras? Even in the same historical period, the susceptibility to offense when it comes to, for example, sexist expression, will vary drastically between cultures and ages—what an American college student might perceive as grossly offensive may be viewed as completely normal by an older person from rural Romania. Instead of trying to objectify the norms for regulating offensive expression, we ought to accept the inherent subjectivity of offense and adjust norms in ways that protects all people from extremely unpleasant situations, regardless of whether their susceptibility is average or not. We ought to be kind and at least to a certain degree make accommodations even for those who do not fit into the norm. Very often people have an

abnormal susceptibility towards offense on certain topics because of the unique hardships they have endured in their lives. Transgender individuals, for example, might be much more upset by someone who mis-genders another individual or someone who argues that we should be free to address people with the pronouns that match their biological sex rather than their chosen ones. Especially when a minority develops a susceptibility to get offended by certain topics, we ought to put their humanity first and treat them with kindness by adjusting the norms of expression in a way that respects their abnormal susceptibilities.

While I believe in the importance of showing kindness and compassion towards the most vulnerable members of our community, I see protecting their right to free expression as much more important for their well-being than shielding them from the discomfort of encountering offense. First, I want to acknowledge that there indeed are challenges with determining what level of susceptibility counts as abnormal. The challenge of defining what is normal, reasonable, and average for legal purposes, however, is overcome by the court judging each case individually against the legal fiction of how the reasonable individual, or in Roman times the *bonus pater familias* would respond if in the shoes of the offended party. Indeed, we cannot define the limits of normal susceptibility to offense on the spot, just as it would be difficult to present an abstract definition of negligence that could be used in all future judgements of negligence cases. However, that does not mean that discounting abnormal susceptibility to offense in cases of offense or looking for abnormally careless behavior in negligence cases cannot produce a just outcome.

While it is extremely important to show kindness and compassion to those who need it most, we cannot accommodate everyone if we want laws to remain just. People with abnormal susceptibilities to offense may indeed have extremely valid reasons for their feelings, but that does not mean that it is the duty of the world to legally recognize their

experience at the cost of other people's fundamental liberties. If we were to make our duty to predict for any possible level of abnormal susceptibility to offense when speaking, we could hardly express anything of significance. The topics we ought to spend time discussing are often the most controversial and challenging ones to discuss. It may be our moral duty to express kindness by offering emotional support to those most affected by difficult subjects and accommodating their sensitivities to an extent, however, we ought not—for their sake as well as ours—shy away from rigorous and daring debate that questions ideas and pushes them to their extremes. The progress of social justice for the most vulnerable individuals after all depends on people's right to use their voice to challenge the status quo. Restricting offensive expression is thus against the interests those who deviate most from the median: adjusting an expression-restricting norm in a way that accounts for the subjectivity of emotion also renders the law much more malleable and thus renders it a tool with which the majority can stifle less popular ideas. While encountering offensive expression may be very upsetting to people, their interests are much more infringed upon and disrespected if their right to free expression is not rigorously protected.

3. Hate

Perhaps the most controversial category of speech that many wish to restrict is hateful expression more commonly referred to as hate speech. Hate speech is notoriously difficult to define but those against it most often describe it as words expressing a speaker's hateful attitude towards the target of his speech and expression which regardless of whether or not it communicates a hateful intention supposedly increases hateful attitudes towards its targets. These two conditions, frustratingly, are neither necessary nor sufficient for defining something as hate speech. Having a hateful attitude is not a sufficient condition for hate

speech—a person can express their hatred without using hateful words by, for example, ironically exclaiming "I just love immigrants. They're leaving such an amazing mark on our country." Neither does intention seem to be a necessary condition for hate speech since people are believed to be engaging in hate speech even when they have good intentions such as for example when they express concern for the target group by saying "being transgender is a mental disorder and we are doing transgender individuals a disservice by playing into their fantasies instead of offering them psychiatric support." Propagating or promoting hate towards a group or a person as a standard poses similar challenges as it is neither a necessary nor a sufficient condition for hate speech: it cannot be necessary or else hurling racial slurs when no one can hear you is not hate speech and it cannot be sufficient since there is expression that increases hatred of a certain group but does not intuitively strike us as hate speech. For example, if an advocate for transgender rights makes a well-meaning argument about transgender athletes deserving to compete in the category of their chosen gender in front of a very conservative audience, he might actually increase the audience's level of hatred for transgender individuals without engaging in hate speech. Despite the difficulties in establishing a definition of hate speech, however, many laws and guidelines have been put forth in a desire to restrict a broad range of expression from "fake" news, speech that is upsetting due to depictions of violence, speech that contributes to "extremist" ideas, speech that mocks religion or "traditional family values," all the way to speech that makes members of a certain group, particularly a minority anxious, uncomfortable, or offended.

In the following pages I will discuss specifically this latter form of "hate speech" that I will for clarity's sake refer to as minority defamation. I will define minority defamation as speech falsely attributing a negative trait to a group of people in the minority. By "falsely attributing" I want to refer to not merely the act of forwarding an entirely false proposition

about a group of people such as “Jewish people are vermin” but also the dissemination of partial truths that are overexaggerated in order to harm a group’s reputation, for example, “Mexicans are criminals”—there certainly exist criminals of Mexican descent but the proposition is a form of minority defamation because it attempts to extend a negative trait true for some to all members of the group. I have decided to discuss minority defamation because the argument for restricting it is common, popular, and seems to be viewed as very persuasive even though it directly clashes with the Severity, Direct Causality, and Last Resort standards that I view as necessary conditions for speech restriction. With the help of Jeremy Waldron’s argument in *The Harm in Hate Speech* I will present a case for restricting speech that does not have a severe negative consequence caused directly by the expression in question that could be addressed by discussion rather than censorship. Because of the negative and unjust long-term state of affairs minority defamation contributes to, legally censoring it is forwarded not only as justifiable but as a moral duty of any fair democratic government.

Expression that may be defined as minority defamation consists of various statements with various levels of objectionability from “Muslims are terrorists” and “transgender people are crazy” to “blonde women are stupid.” Many argue that minority defamation presents a serious threat to the minorities in question. Waldron views the problem of minority defamation as a form of harm deeply problematic because it endangers the minority member’s rights as equal citizens: “The issue is publication and the harm done to individuals and groups through the disfiguring of our social environment by visible, public, and semipermanent announcements to the effect that in the opinion of one group in the community, perhaps the majority, members of another group are not worthy of equal citizenship” (33). The minorities targeted by defamation are, according to Waldron,

condemned to a life of fear: “Can their lives be led, can their children be brought up, can their hopes be maintained, and their worst fears dispelled, in a social environment polluted by these materials?” (33). Waldron’s concerns hint at the following argument:

1. Minority defamation communicates a negative sentiment of a group of people—often the majority.

2. Associating a negative sentiment with a group of people is equivalent to considering said group less deserving of equal citizenship.

3. (from 1 & 2) Minority defamation leads a group of people—often the majority, to consider a minority group less deserving of equal citizenship.

The greatest weakness of this brief argument lies in the second premise. Attributing any negative trait to someone certainly cannot be construed as equal to proposing that they be stripped of their citizenship rights: there are many negative traits such as laziness, stupidity, greed, and promiscuity that have previously been attributed to African-Americans, women, Jewish people, and the members of LGBT community and they have nothing to do with citizenship rights. We may not be fond of lazy, stupid, greedy, and promiscuous members of our society, but these traits do not imply that they ought to be treated differently as citizens. However, there are some characterizations such as being a rapist, a terrorist, a thief, and a crazy person in which there is implicit a suggestion that certain restriction of freedoms are due. Perhaps the idea is that the most pernicious forms of minority defamation are those which make the accusations with the most implications for a different treatment. Nevertheless, a more accurate and more modest second premise, would stop at proposing that associating a negative sentiment with a group of people is equivalent to proposing that they be treated differently. Proposing that a group of people is lazy, stupid, greedy, and

promiscuous indeed implies that they ought to be treated with greater caution and are deserving of less confidence and trust than others.

On both interpretations of the second premise minority defamation is taken to compromise a group's right to equal treatment. Waldron forwards an argument that strives for a more ambitious conclusion, namely, that a minority group's rights to equal citizenship are endangered. However, to an extent his way of reasoning is equally applicable in its more modest formulation which links minority defamation with worse treatment. In both cases, the idea is that a certain genre of expression deprives minorities of an important public good: the assurance that they are equal members of a community, namely "that they can count on being treated justly" (85). Waldron understands one's lack of assurance in the strong sense of the term—as a threat to one's citizenship rights that manifests in one's different treatment by authorities in front of the law. He believes that having Assurance (S) means having a guarantee of being treated equally by the law. With my amendment to earlier premise 2, however, I understand lack of assurance in its weaker sense, namely as a threat to one's equal treatment by one's fellow citizens. I will refer to my usage of the term as assurance (W), defined as having a guarantee of being treated equally by one's fellow citizens.

For Waldron, Assurance (S) is an essential trait of a just society that can be characterized as an underlying confidence of its citizens in the fact that their rights will be respected. He understands preserving Assurance (S) as key to protecting people's dignity in front of the law by ensuring that everyone has the same "social standing" and a "basic reputation" that ensures that people "be treated as equals in the ordinary operations of society" (5) such as the society's legal procedures. Under my view, however, assurance (W) is an essential trait of an ideally just society, not a just society. In the best of all possible worlds, a person must indeed have assurance (W) that no matter whom they encounter, they

will have the same social standing, the same basic reputation, and be treated as equal regardless of his or her belonging to a certain identity group. In reality, however, in-group and out-group ways of thinking, further reinforced by stereotypes, influence the way we respond to and treat other people. While a perfectly just society is indeed one in which every single action of every single person is perfectly just, it is sufficient for a just society that it has laws ensuring the protection of people's rights even if individuals within it treat each other differently based on them belonging to a certain group. In plain terms, in a just society people are free to prefer some groups over others and to an extent display said preferences in their behavior. They may not, however, infringe upon people's fundamental liberties and citizenship rights.

In summary, having Assurance (S) is a necessary condition of a just society as it means knowing that you will be viewed as equal to any other citizen in front of the law. On the other hand, assurance (W) is a necessary condition of a perfectly just society but is not necessary for a just society as it refers to knowing that you will be viewed as equal to any other citizen by any other citizen.

Waldron wishes to argue that minority defamation undermines people's Assurance (S) in two ways. First, he emphasizes that even without infringing on their legal rights, minority defamation strips people of feeling secure in their daily lives due to the threat of being verbally attacked or excluded when they go about their daily business: "A vigilant police force and a Justice Department may still keep people from being attacked or excluded, but they no longer have the benefit of a general and diffuse assurance to this effect, provided and enjoyed as a public good, furnished to all by each" (85). While valid, this consideration does not pertain to Assurance (S) but rather to assurance (w) as it does not bring up a concern regarding people's legal rights being respected. Rather, it discusses their experience of social

ostracization. Second, Waldron asserts that minority defamation provides "a focal point for the proliferation and coordination of the attitudes" behind hateful expression and "a public manifestation of hatred by some people to indicate to others that they are not alone in their racism or bigotry" (95). As such, Waldron understands minority defamation as something that presents a true threat to Assurance (S) as it keeps alive the threat that one day majority members motivated by the associations that they have of minorities with negative sentiments will vote minorities out of their rights.

Motivated by such a concern, Waldron proposes that we impose expression-regulation as a way to prevent the potential aggregation of negative sentiments to the extent that would tip over the public opinion in favor of stripping minorities of their rights. Laws against minority defamation are necessary, according to Waldron, to protect an "environmental good" of allowing minorities to live without fear of systemic oppression and maintaining "a certain ecology of respect, dignity, and assurance" by restricting speech which pollutes such goods (96). He compares regulating group defamation to regulating CO_2 emissions in order to protect the environment: "we figure that the tiny impacts of millions of actions—each apparently inconsiderable in itself—can produce a large-scale toxic effect that, even at the mass level, operates insidiously as a sort of slow acting poison" (97). Even though an individual instance of minority defamation may not clearly cause a severe negative effect, the threat of the aggregate catastrophe just like climate change not only allows but calls for legal action.

Waldron's Assurance (S) Argument may be summarized as follows:

1. Minority defamation allows a group of people—potentially the majority—to proliferate, coordinate, and publicize their negative sentiments of minority members.

2. Associating a negative sentiment with a group of people is equivalent to considering said group less deserving of equal citizenship.

3. (from 1 & 2) Minority defamation allows a group of people—potentially the majority—to proliferate, coordinate, and publicize their view that a minority group is less deserving of equal citizenship.

4. Proliferating, coordinating, and publicizing views that a minority group is less deserving of equal citizenship keeps alive the possibility that the majority will vote in favor of stripping a minority group of equal citizenship.

5. (from 3 & 4) Minority defamation keeps alive the possibility that the majority will vote in favor of stripping a minority group of equal citizenship rights.

6. A just society has a duty to legally restrict anything that endangers an environmental good of all citizens being equal with regards to their rights.

7. Activities that keep alive the possibility that the majority will vote in favor of stripping a minority group of equal citizenship rights endanger the environmental good of equality.

8. (from 6 & 7) A just society has a duty to legally restrict activities that keep alive the possibility that the majority will vote in favor of stripping a minority group of equal citizenship.

9. (from 5 & 8) A just society has a duty to legally restrict minority defamation.

As discussed earlier, Waldron's entire argument relies on a false second premise. Associating a negative trait with a minority group is not equivalent to proposing that they be stripped of equal citizenship. We can, however, drastically restrict the domain of minority defamation Waldron's argument and argue it applies only to those instances of minority defamation in which traits attributed such as criminality and terrorism indeed have a closer

connection to right-restriction. Even if we do so, Waldron's argument faces another fundamental problem with premise #6. A just society certainly has a duty to legally restrict that which compromises the good of all citizens to be equal with regards to their rights. Having a duty to protect an environmental good, however, implies a much greater level of government intrusion. To protect a good of equality, one must prohibit the violations of equality. Protecting an environmental good of equality, on the other hand, requires a prohibition or at least a severe restriction of all the minute things such as actions, attitudes, and thoughts that may in aggregate present a violation. Moreover, according to Waldron, protecting this environmental good justifies restricting free expression—a fundamental human liberty. According to the environmental analogy, what Waldron proposes is therefore comparable not so much to taxing CO_2 emissions as it is comparable to restricting fundamental human liberties such as the right to reproduce and travel around freely due to the contribution said lifestyle choices make to climate change.

I do not wish to challenge Waldron's proposition that a society has a duty to take care of an environmental good of equality. Proposing that it is a duty to restrict actions that contribute to the endangerment of said environmental good, however, is too radical a proposal for as vague a requirement as it is "to contribute" to the pollution of an environmental good. Moreover, placing a duty to protect the environmental good of equality through prohibition of expression that contributes to it is equivalent to saying that it is our duty to restrict CO_2 emissions through travel bans and second child taxes—it is way too narrow. Just as we can address climate change through transitioning to cleaner energy sources, we can address intolerance and inequality through education. We not only can but ought to develop cleaner energy sources and education strategies before we even think about

interfering with people's fundamental liberties such as expression, reproduction, and freedom of movement.

Waldron certainly goes too far both in attributing too great of an effect to expression that speaks pejoratively of a minority group and in attributing to steep a duty to the society in order to remain just. However, perhaps there is a way to salvage the spirit of his argument without saying that "Asians drive poorly" implies that Asian people do not deserve equal rights as citizens and without assuming that it is a duty (or a right) of a government to ensure the protection of vaguely construed environmental goods at the expense of fundamental human rights. To see where a slightly more modest Assurance Argument takes us, I will return to my reformulation of premise #2 based on the weaker sense of the term "assurance."

A less ambitious version of the assurance (W) argument would go as follows:

1. Minority defamation allows a group of people—potentially the majority—to proliferate, coordinate, and publicize their negative sentiments of minority members.
2. Associating a negative sentiment with a group of people is equivalent to proposing that said group of people ought to be treated differently when encountered.
3. (from 1 & 2) Minority defamation allows a group of people—potentially the majority—to proliferate, coordinate, and publicize their views that a minority group ought to be treated differently when encountered.
4. Proliferating, coordinating, and publicizing views that a minority group ought to be treated differently when encountered makes the majority treat the minority group members differently when encountered.
5. (from 3 & 4) Minority defamation makes the majority treat the minority group members differently when encountered.

Thus far, the argument runs relatively smoothly. It is not outrageous to propose that repeatedly attributing a negative attribute to a group leads to some people adopting it as fact and altering their behavior towards the slandered group's members. Proliferation, coordination, and publication of defamatory views further increases the likelihood that majority group members will indeed treat minority members differently because of holding certain beliefs about them. From this moment in the argument, however, we must necessarily change the environmental good that is being endangered from one pertaining to the minority member's rights to one pertaining to them being treated differently by people when going about their daily business.

6. A just society may legally restrict anything that endangers an environmental good of all citizens being treated as equal by their fellow citizens.

7. Activities that make the majority treat the minority group members differently when encountered endanger the environmental good of equal treatment.

8. (from 6 & 7) A society may legally restrict activities that make the majority treat the minority group members differently when encountered.

9. (from 5 & 8) A society may legally restrict minority defamation.

This version of the argument generates a mere permission rather than a duty for a society to legally restrict minority defamation. Such a weaker conclusion seems to co-align with a lot of people's intuitions about restricting these kinds of hate speech. Namely, many believe that if the statements made about a vulnerable group of people are extremely demeaning and spread misconceptions about the minority group that negatively affect people's treatment of its members, they deserve punishment and restriction for making the minority group members' lives unnecessarily unpleasant. The circulation of racist and homophobic stereotypes, for example, according to such a line of reasoning deserves

prohibition because it leads to perfectly legal actions that nevertheless can be viewed as extremely disrespectful such as security guards trailing a black teenager in a clothing store or parents instructing their child not to be friends with a boy adopted by a gay couple.

While superficially alluring, however, this second version of the Assurance Argument faces challenges as well. The remodeling of the second premise might have worked well, but premise #6 forwards a much more radical proposition than it may appear initially. It proposes that there is nothing wrong with a society legally dictating not only that people respect each other's rights but that they respect each other equally in everyday activities. We certainly have a moral obligation to treat our fellow citizens equally regardless of their identity attributes such as religion, race, or skin color. However, it is not the government's place to legally restrict all immoral actions. Treating others with less respect and trust or with greater suspicion and distaste because of their personal circumstance might be extremely morally condemnable, but it ought not be rendered illegal because doing so would violate fundamental liberties of thinking freely and freely associating with people you like.

My objection to premise #6 might appear too cold-hearted. However, I am not saying that the government has no business ensuring its citizens treat each other with respect, merely that they have no business legally micromanaging people's attitudes towards each other—most of all they have no justification legally restricting attitudes at the level of expression and not even action. It is important that the government invests efforts into facilitating social harmony among its citizens, but it must do so by way of education rather than coercion. By educating everyone about the culture and history of different minority groups and presenting the minorities as having an enriching rather than a threatening influence on the dominant culture, the government can encourage respect without forcing it upon its citizens which is often more effective in the long term. The environmental good of

equal treatment on an intrapersonal level is not something that a society has by default and is taken away from it through small actions such as minority defamation. Rather, ensuring social cohesion is something a society must actively work for by finding inventive ways to persuade its members to learn about different cultures and view them as fellow citizens co-creating their culture rather than foreign forces threatening it.

4. The Bane of Speech Regulation

Discussing the regulation of speech that leads to harm, offense, and hate brought to light a recurring pattern—when viewed in abstract, the three regulatory standards that I propose appear grossly insufficient for capturing all expression that we feel ought to be punished and restricted. Permitting the government to meddle only with expression that directly causes severe harm or offense that cannot be prevented by other means than censorship seems to leave too many people unprotected from harmful, upsetting, and unpleasant expression of others. Furthermore, because hatred for minorities is never a direct cause of any given expression but rather forms gradually and overtime with many contributing factors, my proposed regulatory standards leave the most vulnerable of our society unprotected by law against the soiling of an environmental good that most people can take for granted. In abstract, Severity, Direct Causality, and Last Resort standards seem way too narrow.

Every attempt to expand them, however, poses a great threat to inalienable human liberties and important societal goods. When it comes to harm, allowing for the censorship of expression that leads to less severe, and less directly caused harm renders speech-regulation so much more subjective that it allows for political abuses such as we have seen in the application of the Bad Tendency test. Legally protecting people from more harm caused by expression thus comes at the cost of their right to be active politically and express

controversial views. Similarly, censoring offensive expression in a way that fully protects people from extreme discomfort and upset requires rendering regulatory standards so subjective that just about any opinion about public policy, personal values, or morality may be prohibited because it is sure to offend someone. Legally protecting people from more than just the most severe and outrageous offense thus comes at the cost of their right to express their thoughts about anything that someone may strongly disagree with. Lastly, censoring expression that shines a negative light on a minority group disallows the expression of discontent with another group's role in the society, judgements of other's culture or way of life, and the voicing of personal preferences. While expressing personal biases about a minority group may not be morally praiseworthy, people's ability to do so plays an important role in a multicultural democratic society because it exposes the ways in which we are failing to facilitate group cohesion and allows us to address our shortcomings by publicly correcting biased views. Legally protecting minorities from hateful expression thus prevents us from identifying areas in which efforts to facilitate social harmony and respect are most needed as well as impedes on understanding the reasons for intragroup hatred.

The bane of regulating expression, therefore, lies in the fact that the conditions that appear to narrow down the pool of expression subject to restriction too much are in fact the only ones that prevent overregulation at the cost of fundamental human liberties and important societal goods that help us forge better democracies and more tolerant and respectful communities. Setting minimalistic standards for expression-regulation violates popular intuitions so much precisely because it is unprecedented and counterintuitive to value other people's right to free expression. In a 1992 documentary *Manufacturing Consent: Noah Chomsky and the Media*, Chomsky points out that "if you're really in favor of free speech, then you're in favor of freedom of speech for precisely the views you despise." Our

completely natural tendency, however, is to wish for our own freedom of speech but not for the freedom of those whose opinions we find most abhorrent. Only by sticking to the Severity, Direct Cause, and Last Resort standards, we can restrict expression without sacrificing the benefits of free expression for personal and public intellectual life.

5. The Importance of Minimal Regulation

Why would an experienced police officer need to bother with procuring a warrant before searching a suspect's property when the suspect all but admits to storing the murder weapon in their bedroom? Why should secret service not gain access to information exchanged through end-to-end encrypted messaging platforms if they know them to be a tool of organizing terror attacks? And why should we not silence an opinion of a vaccine hesitant public figure if we know they might lead people not to take the Covid-19 vaccine? It is because all persons have fundamental human liberties such as the right to due process, privacy, and freedom of expression. Such individual rights are frustrating for law enforcers, lawmakers, and governments alike. Even regular citizens often see violating them as justifiable in some cases and view having to respect them as an occasional annoying obstacle to a better society. To fight against relativistic tendencies to bend principles defending fundamental human liberties whenever convenient for us, I put my best efforts forth to provide a comprehensive defense of freedom of speech.

As early as in Ancient Greece, freedom of speech was believed to have a positive impact on the accuracy of Athenian communal decision-making, it instilled courage into Athenians when going to war, and gave them a sense of self-determination. In Locke's time, freedom of expression was recognized for its role in allowing for freedom of thought and belief. Some viewed it as important to the protection of human dignity as the right to private property. Mill reintroduces and upgrades Lockean fundamentalist view of freedom of

speech's importance for freedom of thought by armoring it with further epistemic arguments for the importance of free expression in truth seeking. Courage, autonomy, and truth are only few of the many values we infringe upon when overzealously restricting free expression. In light of the historically recognized significance of freedom of speech, the optimal way of regulating speech is to only censor speech that satisfies the Severity, Direct Cause, and Last Resort standard. Admittedly, there are situations in which I recognize the allure of bending the standards for the sake of satisfying our intuitions, but I invite us to weigh the benefit of preventing harm or offense or hatred in a specific circumstance which does not satisfy the three standards against the harms brought about by expanding the pool of expression that the government can infringe upon. Just as it is not worth it to abandon the due process and our messaging privacy to catch a couple of crooks and terrorists faster, it is not worth removing some harm, offense, and hatred from circulation if removing it opens the gateway to censorship of information, opinion, and ultimately thought.

C. Science and Hate Speech Regulation

1. Popular Opinion: The Neuroscience of Hate Speech

Even though neuroscience has for long studied mental phenomena related to hate speech such as empathy, prejudice, and distrust, data from cognitive sciences has relatively recently gained attention in the discourse about hate speech regulation. A 2018 NYT opinion piece "The Neuroscience of Hate Speech" by a psychiatrist Richard A. Friedman encapsulates most widely shared opinions about the "science of hate speech" that have been even more frequently circulated during Trump's presidency. Friedman expresses the view that it is obvious that speech that denigrates and misrepresents a group has a connection to the frequency of hate crimes: "You don't need to be a psychiatrist to understand that the kind of hate and fear-mongering that is the stock-in-trade of Mr. Trump and his enablers can goad deranged people to action." He presents several assertions as facts of science such as that hate speech increases prejudice, provides a surge of stress hormones that make it hard for people to think before they act, and dehumanizes the target group which results in the members becoming victims of hate crime. Friedman thus argues that according to science, hate speech, especially from a figure of authority, causally contributes to hate crimes as he dramatically concludes "Now imagine what would happen if President Trump actually issued a call to arms to his supporters. Scared? You should be." I take Friedman's column to be a good summary of the subjective and often flawed interpretations of data that, especially in the Trump era, many have adopted as an objective fact. Science to an extent establishes a correlation between hate speech and hate acts, but said correlation could be mediated by a number of factors and is by no means sufficient to establish causation. Nevertheless, Friedman-like thinkers have been over-confidently presenting their interpretation of studies on hate speech as grounds for speech-restricting policy.

The propositions of Friedman's column are popular and frequently expressed but as far as science is concerned at best only one of the possible interpretations of the data and at worst a plain misreading of science. Let us, for example, examine Friedman's first proposition: "repeated exposure to hate speech can increase prejudice." Friedman states this as a plain fact and refers to a series of Polish studies from 2017 for support. The study, consisting of three separate experiments, however, paints a much more nuanced picture of hate speech's relation to prejudice. The first experiment measures the correlation between self-reported exposure to hate speech and prejudice assessed by asking the participants to rate their preferred level of distance to outgroup members: the lower the distance the less opposed the participant is to having the outgroup member as a friend, neighbor, colleague, etc. After analyzing the data, the scientists "did not observe any evidence of a simple relation between frequency of exposure to hate speech and prejudice" (Soral et al, 4). More surprisingly perhaps, those with more exposure to hate speech demonstrated lower prejudice as they "were preferring lower distance towards outgroup members" (Soral et al, 4). One of the possible interpretations of the latter finding is that "contact with hate speech might also have some positive consequences (e.g., raised compassion for the victims)" (Soral et al, 4). The second experiment aimed to establish that exposing a participant to hate speech on the spot in an attempt to desensitize him to it increases the participant's outgroup prejudice. The scientists indeed confirmed their finding (which is probably what Friedman refers to) but found no difference between the desensitized and not-desensitized group after accounting for the participant's prior level of sensitivity to hate speech: "after controlling for the level of sensitivity to hate speech, the estimated difference reduced to nonsignificant, while the level of sensitivity to hate speech significantly predicted outgroup prejudice" (Soral et al, 6). In other words, the correlation between hate speech exposure and prejudice turned out to be

spurious after controlling for the level of sensitivity to hate speech. The group reports a similar phenomenon in their final experiment which replicates study 2 with anti-immigrant attitudes rather than general prejudice: "after accounting for both mediators [sensitivity to social norms and sensitivity to hate speech], the direct effects of exposure to hate speech on outgroup prejudice and anti-immigrant attitudes became reversed or non-significant, respectively" (Soral et al, 8). In brief, the study fails to establish that higher exposure to hate speech rather than the participants' prior traits is a good indicator of outgroup prejudice. While the study offers interesting insight, it is not what Freidman presents it as—a scientific proof that hate speech exposure can increase prejudice with unsaid but heavily hinted at implications for the need for censorship.

Another concerning misinterpretation of science occurs when Friedman suggest that public hate and fear mongering can cause violence by triggering a release of stress hormones in people's amygdala: "politicians like Mr. Trump who stoke anger and fear in their supporters provoke a surge of stress hormones /…/ and engage the amygdala /…/ This makes it hard for people to dial down their emotions and think before they act." While the study Friedman refers to prefaces is finding with the statement that "direct evidence for [the amygdala's] role in the emotional processing of linguistic cues is lacking," it does suggest that the activity of the amygdala—a part of the brain responsible for releasing stress hormones such as norepinephrine and cortisol—correlates with threatening language. While correlation is not the same as causation, Friedman, to give him the benefit of the doubt, might be hypothesizing causation because of his familiarity with a great number of studies performed on animals that make a causal connection between stress and amygdala activation very likely. However, Friedman at best oversimplifies and at worst misconstrues the mechanisms related to decision-making when he suggests that politicians might have caused

violence by triggering people's amygdala with their fearmongering speech: "Mr. Trump has managed to convince his supporters that /…/ we face an existential threat from imagined dangers /…/ Where the men arrested in the synagogue shootings and bombing attacks listening?". Committing a violent act, especially a premeditated one like a shooting spree or a bombing requires much more complex brain activity than simple amygdala firing. A person's amygdala fires up countless of times a day in response to as little as a light flashing suddenly, or even just seeing color red. If stress hormones released by the amygdala were as powerful and as overwhelming as Friedman suggests, it would be dangerous to even conduct studies in which scientists evoke amygdala activity. Prefrontal cortex is the center of decision-making. Any individual without a severe impairment of their prefrontal cortex makes decisions not only according to physiological cues like levels of blood sugar and stress hormones but also according to his beliefs and preconceptions. Suggesting that a politician may be partially responsible for someone's crimes because his speech triggered someone's amygdala, disregards the complexity of human cognition and unfairly relieves the culprits of the full responsibility for their actions.

The last debatable proposition Friedman forwards is that hate speech leads to dehumanization which leads to people treating the dehumanized group more poorly: "when someone like President Trump dehumanizes his adversaries, he could be putting them beyond the reach of empathy, stripping them of moral protection and making it easier to harm them." Yet again, however, Friedman puts forth a much stronger conclusion than the one suggested by the Harris and Fiske study from 2011 that he links. Harris and Fiske establish firstly, that when tasked with describing first a day in the life of a typical person and then of a homeless person, the participants use comparatively fewer words that signal that they are imagining the homeless person's state of mind and secondly, that the participants' anterior insula, a region

of the brain implicated in disgust, interception, and pain and punishment neural network, activates less when the participants view a picture of a member of a social group that they rate highly in warmth. In suggesting that the study establishes that we are more eager to harm the targets of dehumanizing speech, Friedman presents only one of the several possible interpretations of the data. The results of Harris and Fiske's first finding could be explained by the fact that we have a harder time imagining the lives of people whose way of life resembles ours less, whereas the second study merely establishes that human-perception dimensions correlate with brain regions connected to disgust. Just as amygdala firing does not guarantee violence, disgust neural networks activation is not exclusive to dehumanization, let alone dehumanizing conduct. For lack of a less overused phrase, correlation does not equal causation.

Contrary to Friedman's suggestion, studies have found that dehumanization is not exclusive to intergroup violence. Most notably, Simon and Gutsell in a 2021 study show that dehumanization, defined as a failure to recognize cognitive and emotional complexities of the people around us, occurs in completely everyday activities—as far as our brain activity is concerned, we do not treat people as equally human in our everyday interactions. Unsurprisingly, however, the fact that our brain has a hard time recognizing someone's cognitive and emotional complexity does not mean we have a sudden urge to harm them. Not all dehumanization is malicious and leads to actions of violence. In their 2016 paper Cameron, Harris, and Payne suggests that dehumanization at times occurs in order to reduce emotional costs involved in helping stigmatized groups. In short, anyone with an understanding of the scientific method and capable of reading and interpreting studies will see that the connection between dehumanization, hate speech, and hate crimes is extremely complex and context dependent. Friedman in his opinion piece justly calls for political

leaders to think more about the effects their words may have on people and the way certain kinds of language can feed pre-existing biases. It is fair to say that politicians affect public opinion. However, it is not correct to assert that science shows that those spouting hate speech are responsible for hateful actions or even hateful sentiments of people listening. Responsibility requires causality that cannot be established by appealing to a few correlations–especially not when we are dealing with a complex and messy web of factors that contribute to the creation of group hatred. While Friedman's call for greater civility is well placed, it cannot claim the high ground of an empirical, scientific truth.

2. Dehumanization and Hate Speech

While Friedman poorly supports the idea that hate speech causes hate acts, many find the connection intuitively appealing. Because people often turn to science to prove it, I will discuss another attempt to integrate findings of neuroscience to extrapolate conclusions about the expression and the enactment of hatred.

In a 2015 paper Gail and Richard Murrow attempt to explain what role dehumanization has in hate acts. They propose that dehumanization "has an automatic dampening effect on the neural mechanisms of pain empathy that enable empathy for the pain and suffering of others" and propose that the neural mechanisms of pain empathy of someone who dehumanizes a group of people by associating them with subhuman traits "do not respond to the pain or suffering of that dehumanized category as robustly" which may explain why they are more likely to hurt people they dehumanize (Murrow and Murrow, 337). They see their theory as having potential implications for First Amendment jurisprudence as well as for the dangers of permitting the expression of dehumanizing hate speech.

The first part of the Murrows' hypothesis aims to causally connect dehumanization with lack of empathy, but it relies on the (at best) questionable assumption that empathy is a

conspecific phenomenon, namely that we empathize only with pain and emotion of another human. The Marrow's need this to be true because they are trying to establish that we lose empathy for people because we consider them less than people. Such a claim goes against a body of research that came to light since Premack's seminal findings which showed that children recognize complex properties and behaviors even when represented with geometrical forms (Premack, 1990). Admittedly, in animal studies researchers often note greater or exclusive responsiveness of subjects to the behavior conspecifics (Nieuwburg, 2021). However, our ability to empathize with animals as well as our capacities for abstract reasoning and emoting suggest that "our brains come pre-equipped with a broadly generalized, and not at all conspecific, capacity for empathy" (Hoffman, 168). The Murrows' suggestion that dehumanizing someone causes a decrease of empathy is thus ill supported. A decrease of empathic response may cause dehumanization, or the two phenomena may be only correlated.

The Murrows, however, continue to assume a causal connection between dehumanization and lack of empathy and try to explain it through postulating implications of research in 'mirror neurons.' Mirror neurons are most robustly studied in primates who have been observed to experience an automatic mirror neural simulation when observing others perform certain motor acts. While they were discovered in the context of motor research, however, the idea of mirroring has been applied to the study of empathy for pain: "Though much neuroscientific research has looked at pain empathy, the ground-breaking study was an *f*MRI study conducted by Singer et al., in 2004, which provided early evidence that when one individual observed a 'sign' that another individual was in pain, some of the same neural components that are active in the firsthand experience of pain were also active in observing or imagining another person in pain" (Murrow and Murrow, 346). However, as presented in

The Student's Guide to Social Neuroscience, the idea that emotion imitation equals empathy "is an over-simplification" and "the link between mirror neurons themselves and imitation is by no means uncontroversial" since "monkeys (who posses mirror neurons) have very limited imitation ability" (Ward, 180). In other words, to the extent we currently understand mirror neurons, a mirroring firing of neural pathways shows at best that you recognize an action or an emotion; it by no means establishes that you will imitate it or are capable of doing so.

The Murrows have to rely on numerous dubious or even false assumptions in order to put forth their dehumanization=lack of empathy=right-violating conduct hypothesis. However, they go as far as to suggest that if accepted, their hypothesis bears implications to hate speech as it "may shed a light not just on how hate speech might *enable* or *incite* violence, but how it can do so in such an apparently unthinking, 'banal', perfunctory, *rapid* manner" (Murrow and Murrow, 354). They believe that their theory explains how "hate speech goes *around* the *conscious mind* to directly attack the *emotional* mechanisms of empathy or moral restraint" (Murrow and Murrow, 355). However, they fail to provide evidence for hate speech causing the dehumanization. It is ludicrous to suggest that anyone who hears a hateful comment has suddenly had their conscious mind breached and their emotional mechanisms for empathy impaired. The Murrows' article explores how dehumanization relates to empathy, but to show that hate speech causes dehumanization and leads to hate crimes, they would need to conduct experiments to show that people's emotional mechanism of empathy towards a group falter after exposure to hate speech and their proclivity to harm members of the group increases.

Mirror neuron theories and studies mapping brain activity, in the way they are currently conducted, do not carry the answers to these questions. Even if we assume, without clear evidence, that hate speech triggers neural networks related to hateful actions, it may do

so because our brains recognize the content of the speech and not because we are inclined to imitate it. Roginsky and Tsesis succinctly summarize the problematic way in which the Murrows handle scientific evidence: “connecting the scientific concept of mirror neurons as a potential explanation for individuals’ capacities to participate in discrimination, and worse, is too much of an extrapolation. The original authors of the scientific papers on mirror neurons described them as part of a motor function process, not a theory to help explain the darkest times of history and human behavior” (Roginsky and Tsesis, 177). To their credit, the Murrows acknowledge at several points that they are forwarding mere hypotheses and do not deem their theory concrete enough to ground policy change. But to reach their conclusion they need to stack suppositions and postulate causation where proof offers mere correlation to the extent that renders their theory not only inapplicable to policy but also of very limited scientific value. Societal hatred is an extremely complex social phenomenon. Neuroscience can offer us insight into the mental processes that accompany hate. It can point at parts of the brain that activate in response to certain stimuli. While it can tell us what happens in our brain when we discuss or experience hatred, however, it is far from being able to help us identify and prevent the causal process behind hate crimes.

PART 3: WHAT MOTIVATES OVER-REGULATION OF FREE EXPRESSION?

A. The "Speech = Violence" Fallacy

There has undeniably been a shift in young people's sentiments regarding speech-regulation. Whereas liberal youth in the first half of the 20th century as well as in the Vietnam War era used to be at the frontlines of the battle defending the freedom to express controversial opinions such as anti-war, pacifist, and communist sentiments, young people nowadays more frequently advocate for punishing or restricting speech they view as offensive, triggering, or upsetting. The shift in young people's attitudes towards free speech may be explained by examining the shift in a key fundamental assumption at the background of discussing speech regulation—whereas we once held speech to be the antidote to violence, it is now more and more commonly perceived as violence itself. With the aid of French philosopher Paul Ricoeur, I will examine the philosophical ideas that facilitated the understanding of speech or language as violence. While I acknowledge that to an extent it is possible in figurative terms to discuss language as violence, I will take a strong stance against the proposition that speech be understood and treated literally as violence in the everyday functioning of our society. After presenting and opposing the idea that speech equals violence from a theoretical standpoint, I will examine a psychologist's argument that aims to employ findings from psychology to justify treating some speech as violence. A more careful discussion of the science employed will reveal that equating speech with violence is neither theoretically nor practically plausible.

1. Speech as the Opposite of Violence

In 1893, Sigmund Freud co-authored an academic paper in which he conveyed the idea that "the man who first flung a word of abuse at his enemy instead of a spear was the

founder of civilisation."[4] While this first catch-phrase is well known and frequently cited, Freud continues his sentence by making a point about the relationship between words and actions: "thus word is a substitute for deed and in some circumstances the only substitute."[5] Freud's statements emphasize that the evolution of language enabled us to use speech instead of action to communicate negative emotions. Such a shift from actions to words turned out to be a ground-breaking invention that enabled the development of modern human society—specifically because it promoted non-violent ways of resolving conflict. Anyone even vaguely familiar with human history may rightfully doubt the idea that there has been less violence amongst the human species since we have developed the capacity for speech. Indeed, the development of speech enabled mankind to practice new and often even more pernicious forms of violence such as wars, organized crime, genocide and so forth. However, whereas in a world without speech the only way to express dissatisfaction or disagreement with a fellow human was to act negative feelings out through violent outbursts, a world with speech gives us a way to air our grievances in a non-violent way and develop laws and contracts that disincentivize members of our modern societies to engage in violence. Violence and speech ought therefore be understood as two possible ways to achieve an end. In a society prioritizing civilized and constructive conflict-resolution speech rather than violence is the go-to way of attaining goals. Violence replaces speech only when the situation calls for more than "mere words" and requires a use of force.

In the philosophical tradition violence and speech are often presented in opposition to each other. For example, Paul Ricoeur in "Violence and Language" argues that only once we gained the capacity for speech, we began viewing violence as a negative tool to achieve one's

[4] **"**derjenige, welcher dem Feinde statt des Pfeiles ein Schimpfwort entgegenschleuderte, war der Begründer der Civilisation" (Breuer und Freud, 1893)
[5] "…so ist das Wort der Ersatz für die That und unter Umständen der einzige Ersatz" (Breuer und Freud, 1893)

ends: “It is for a being who speaks, who in speaking pursues meaning, who has already entered the discussion and who knows something about rationality that violence is or becomes a problem” (33). Ricoeur thus presents violence and discourse as opposing concepts that define each other through their contrast. Namely, just as language renders violence a condemnable way of achieving goals, our linguistic capacities derive an aspect of their sense from being non-violent: “Speech, discussion, and rationality also draw their unity of meaning from the fact that they are an attempt to reduce violence” (33). He does not deny that speech can be used to promulgate violence but presents pro-violent speech as an inherently contradictory phenomenon: “a person cannot argue for violence without contradicting himself, since by so arguing he wants to be right and already enters the field of speech and of discussion, leaving his weapon at the door” (33). In other words, while speaking and arguing, a pro-violent individual tries to achieve his goal not through violence but through discourse thus undermining his own idea that violence rather than speech is the way to achieve his goals. Ricoeur notes that Eric Weil in *La Logique de la philosophie* occupies a similar position by presenting “coherent discourse and violence” as opposites (33). It is difficult to imagine how two phenomena that through opposition define each other may be possibly understood as one and the same.

2. Speech as Violence

And yet speech and violence stand in a more complicated relationship than that of a clear opposition. In Ricoeur’s words, “violence speaks” (34). To examine the sense in which speech is violence while despite being commonly construed as its opposite, Ricoeur walks us through three different kinds of speech in which it is possible to equate violence to speech itself: political, poetic, and philosophical.

The kind of political speech Ricoeur views as a good example of violence is political discourse conducted within a tyranny. According to Ricoeur "philosophy denounces tyranny precisely because it invades philosophy's territory: language" and "makes its way by seduction, persuasion, and flattery" (35). To exemplify the proposition that this genre of speech equals violence Ricoeur points out that in Nazi Germany, a sophist like Goebbels needed to manufacture expressions "that mobilize hate, that consolidate the society of crime, and that issue the summons to sacrifice and to death" (35). In other words, political communication in tyranny is violence because it manufactures the conditions necessary for violence to occur. Apart from tyranny, Ricoeur presents laws as a genre of violence because they spring from an individuals' decisions to give up their private violence to control everyone through a great, looming, communal threat of violence: "the rule of law which gives form to the social body is also power, an enormous violence which elbows its way through our private violence" (36). One can indeed interpret legal communications and tyranny's propaganda as violence. But such an interpretation calls for an argument as to why we would assume the stronger proposition, namely, that these two forms of speech are equal to violence, rather than the more moderate one, namely, that these forms of speech perpetuate or enable or bring about violence.

Similar to political speech, some aspects of poetic speech may be viewed as a form of violence. To explain the rationale behind viewing poetry as violence, Ricoeur leans heavily onto Heideggerian philosophy that understands meaning-creation as a process of drawing Being into the open. Speech, as Ricoeur interprets Heidegger may be viewed "as the submission to the prescriptions of a measure of Being to which man is originally open" (36) which in simpler terms means that speech unravels the aspect of reality that has a potential to be understood through the medium of the speaker. Such channeling of reality is non-violent

in its core and yet is violence because of the Heideggerian perception of a nature of a word. The word according to Heidegger "establishes a being in its Being and thus preserves it in its openness" (37). If we metaphorically compare openness to a clearing in a dark forest from the depths of which Being is dragged, a word forcefully pins down Being and preserves it in the clear. Under such an interpretation of reality, the Heideggerian poet engages in violence insofar as he enables the "poetic *abduction*" (37) of meaning: "The poet is the violent man who forces things to speak" (37). Even though the complexity of Heideggerian philosophy likely eludes a simple reduction of its propositions into analytical terms, such an elaborate description of the purported violence of poetry begs many questions. For one, the Heideggerian view seems perfectly consistent with the proposition that any meaning-creation that brings Being into the open i.e. exposes some true or real aspect of our existence is a violent act. Therefore, at least this simplified version of what may very well be a more elaborate argument gives little reason for viewing merely poetic speech as well as merely speech rather than other forms of meaning-creation as violence.

Apart from political and poetical language, Ricoeur presents philosophical discourse as one fundamentally interwoven with violence. Firstly, he presents the specificity of an initial question of a philosophic inquiry as a violent one: "To begin is always an exercise of force, even and especially when one begins with absolute substance, as does Spinoza" (37). In other words, the very process of beginning a philosophic inquiry—especially by setting a course for the discussion with axioms as Spinoza does—may be perceived as a forceful and hence, violent action. Secondly, Ricoeur presents the specificity of a course of thinking inherent to doing philosophy as a violent phenomenon because it relies on assumptions or presuppositions (37). Assumptions such as Spinoza's axioms are necessary to begin a philosophical conversation, but they do force the inquiry to take a certain direction in a way

that Ricoeur calls violent. Lastly, philosophical language is burdened by "the violence of the always premature conclusion" (38). Any course of philosophical thinking is necessarily interrupted by practical matters. For example, publishers or journal editors want material of a certain length while philosophers end their conversations because a class or meeting runs to an end or an errand calls them away. Indeed, we can refer to philosophizing an act of violence because of such traits constraining philosophical discourse. However, just like in the case of poetic language, it is unclear that these traits are exclusive to philosophical conversation rather than a trait of all communication which also often needs to be cut short, bases itself on frequently arbitrary assumptions, and may begin somewhat forcefully.

All three of Ricoeur's examples of speech that equals violence share a problem—they propose to stretch the definition of violence so widely that almost nothing remains outside of it. Leibsch in a contemporary article characterizes this trait of the attempts to reveal a deeper relationship between speech and violence as follows: "the suspicion arises that violence contaminates language as a whole, it soaks through everything that demands to be said, and it threatens to affect all speaking, acting and thinking" (11). Tearing down the opposition between discussion and violence that defines violence as problematic destroys the theoretical foundation for rejecting violence and morally condemning it: "violence can no longer simply be seen as a pathological exception, interruption or cancellation of an initially unaffected normality. It rather appears to be normal to expect violence virtually anywhere" (11). It thus appears that a key battle is fought in the field of defining language and violence: either we permit violence's definition to subsume all kinds of language and lose the ability to morally condemn it, or we morally condemn it and restrict its definition in perhaps a slightly stricter way than some people's intuitions desire.

Ricoeur recognizes this fundamental paradox of delineating between language and violence and proposes that we "hold it to be a formal, though empty, truth, that discourse and violence are the most fundamental opposites of human existence" (40). In other words, we can technically use term violence to describe various aspects of expression from creating meaning through denomination to rupturing a conversation. Nevertheless, we really should not extend the definition lest we risk not being able or willing to morally condemn violence since "he who has never ceased pointing violence out as the opposite of discourse will be forever safeguarded against being its apologist, against disguising it, and against believing it superseded when it has not been" (40). Rejecting the proposition that speech is violence thus proves itself imperative to the rejection of violence and the protection of the most universal moral maxim of them all—thou shalt not kill. As we shall see in the later discussion of the current state of academia, the correlation between the rising rates of students who sympathize with the idea that violence is a justifiable way to silence speech and the rising rates of violence-tolerant viewpoints and violent protests on college campuses suggests that those eager to equate speech with violence may very well be the ones most eager employ violence to silence speech.

3. "Violent" Speech

Even though there might be a theoretical incentive to distinguish between speech and violence, we encounter speech in our everyday life that seems as if it really could be called violence because of the effect it has on us. Harassment, threats, and intimidation in particular can put our bodies in a state of psychological stress with very real physical manifestations. There are scientific discoveries regarding the connection between psychological and physical ills that have been employed in the battle to define at least some kinds of speech as violence. To examine the idea that we ought to reclassify some speech as violence out of respect for

science, I will discuss an argument made by a distinguished professor of psychology from Northeastern University Lisa Feldman Barrett in an opinion piece for The New York Times and a critique of Feldman Barrett's opinion as put forth by Pamela Paresky for Psychology Today.

In "When Is Speech Violence?" Feldman Barrett argues that some forms of speech, particularly "speech that bullies and torments" (Feldman Barrett), can from the perspective of our bodily reactions to them be called forms of violence. She begins her argument by listing a couple of scientific findings about the physiological effects of chronic stress: first, chronic stress is one of the conditions under which our immune system's own proinflammatory proteins cytokines spike and may cause physical illness[6] and second, chronic stress accelerates the shortening of telomeres, the length of which is indicative of one's longevity.[7] According to Feldman Barrett, the connection between words and violence is fairly straightforward: "If words can cause stress, and if prolonged stress can cause physical harm, then it seems that speech /…/ can be a form of violence" (Feldman Barrett). Feldman Barrett acknowledges the importance of not preventing beneficial short term-stress that certain educational activities such as engaging with offensive arguments might evoke. Only chronic stress warrants cancellation in her opinion, namely "the kind of stress that brings on illness and remodels your brain" which she understands to be caused by exposure to "rampant bullying" or to a "constant, causal brutality" (Feldman Barrett). Defining speech as a form of violence would therefore be permissible only when said speech evokes sustained levels of stress severe enough to cause physiological changes.

[6] For an interesting meta-analysis of this relationship see Rohleder's "Stress and inflammation—The need to address the gap in the transition between acute and chronic stress effects" (2019).
[7] For a meta-analysis of the complex relationship between the two see Jiang et al.'s "Basal cortisol, cortisol reactivity, and telomere length: A systematic review and meta-analysis" (2019).

Up to this point Feldman Barrett's argument offers itself to little critique as it leans on an idea strongly supported by science: that which causes chronic stress may cause in us the same kinds of physiological changes that we associate with violence. However, Feldman Barrett goes on to argue that said scientific reality makes it "reasonable, scientifically speaking, not to allow a provocateur and hatemonger like Milo Yiannopoulos to speak at your school" (Feldman Barrett). This policy proposal indeed seems to contradict her earlier acknowledgement that short-term stress of educational activities can be beneficial and ought not be prohibited. The contradiction is due to the fact that she does not view the presence of speakers such as Yiannopolous as temporary, but rather presents him as "something noxious, a campaign of abuse" and emphasizes that the discursive value of his speech is low— "debate is not what he is offering" (Feldman Barrett). There is merit to Feldman Barret's idea that universities need not and perhaps should not waste their resources on speakers whose focus is not to illuminate their listeners through intellectually rigorous discourse. However, Feldman Barrett's more ambitious conclusion, namely that speakers like Yiannopoulos ought not be allowed on campuses because their speech is violence, seems highly contestable, least of all because it does not follow from the science Feldman Barrett offers in its support.

In her Psychology Today piece, Paresky's main objection to Feldman Barrett's interpretation of science is that it oversimplifies the process by which speech can cause physical damage on us: "Feldman Barrett fails to consider several mediating factors in the supposed causal chain between words and deleterious physiological effects.". In particular, Feldman Barrett fails to discuss the difference between stress and perceived stress and how the latter is the only good indicator of what physiological effects an environmental factor will have on us. When it comes to controversial speakers, Paresky thus emphasizes that the way we perceive the speaker will have a much greater effect on our reaction to him than his actual

offensiveness level: "If one person tells herself that listening to a speaker is going to be intolerable and harmful, it stands to reason that the experience will be more stressful for her than it will be for the person who tells herself it will be illuminating, or an opportunity to defeat a bad idea." The way we think about things influences their physiological impact on us: if we believe particular stimuli or our lives in general are stressful, we will likely suffer more negative impacts from being exposed to them.[8]

Because it is often our beliefs about our environments rather than our environments themselves that hurt us, mindfulness training which focuses on addressing our perceptions of stress rather than external stressors has been shown to have a beneficial effect on stress and anxiety level as well as on sleep quality and general wellbeing.[9] Thus, Paresky warns that not only is it not true that some speakers enact violence upon their listeners, it is harmful to the students to tell them so: "Students who believe that hearing certain words or listening to certain speakers can harm them may, in fact, succumb to a self-fulfilling prophecy" (Paresky). What any responsible educator on college campuses ought to do instead is to encourage students to change their ways of perceiving the speakers whose ideas they find disagreeable. Just like with mindfulness training, shifting interpretations of presences such as Yiannopoulos may "counteract both the purportedly malignant ideas, and whatever harm might otherwise result from potential stress of listening to them" (Paresky). According to Paresky, speech may never be scientifically viewed as violence—it may lead to violence-like effects but only if we choose to interpret it as such.

---

[8] Perceived stress has been associated with an increased risk of events such as stroke (as an example of a meta-analysis see Booth et al "Evidence of perceived psychosocial stress as a risk factor for stroke in adults: a meta-analysis" (2015)). Displaying a similar pattern, perceived racism has been linked to more adverse mental health (for a meta-analysis see Pieterse et al "Perceived racism and mental health among Black American adults: A meta-analytic review" (2012)).

[9] See Bartlett et al "A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials" (2019)

Paresky is right to point out the importance of perceived stress and to object to shunning speakers with controversial ideas from college campuses because of the harm their words may cause. However, it might be too strong to claim that speech can never be legitimately viewed and treated as at least equivalent to violence. In situations in which verbal abuse is severe, targeted on an individual, and goes on for prolonged periods of time, speech may indeed be considered as verbal violence—because it is violent. It brings about physiological and psychological consequences that often manifests in bullying victims in ways similar to the victims who suffer from physical attacks. Contrary to Feldman Barrett's implications, however, such violent speech has already been recognized as illegal and has been prosecuted under harassment laws while its seemingly more innocent schoolyard versions are tirelessly fought through anti-bullying incentives. What Paresky picks up upon in her argument and what Feldman Barrett overlooks, is that those arguing that speech is violence have long since lost interest in verbal harassment. Rather, they wish to cherry-pick scientific evidence backing the harmfulness of verbal abuse and apply it more broadly so as to get an excuse to censor expression and ideas for what appear to be political rather than scientific purposes.

PART 4: THE DANGERS OF ENDORSING "SPEECH=VIOLENCE" FALLACY

A. Trigger Warnings

1. The Idea

The notion that speech equals violence has taken many forms in the past decades. One of its most successful yet controversial intellectual descendants is the idea that it is right and perhaps even necessary for instructors to give trigger warnings before their students engage with potentially upsetting material. Some claim that the term "trigger warning" applies to certain common practices such as displaying the MPA rating of a movie before showing it. However, unlike the MPA rating that is imagined as a tool for a viewer to decide whether or not to watch movies with younger people around, trigger warnings in the classroom context are declarations that context may be disturbing or upsetting intended to help students mentally prepare for viewing the content or even avoid it if so desired. In their narrowest scope, trigger warnings are aimed at a very small subset of the population: the victims of trauma such as war veterans or rape survivors who suffer from Post-Traumatic Stress Disorder (PTSD) and could potentially re-experience their trauma if exposed to "triggering" content. However, since their humble beginnings, the mission of trigger warnings has expanded to protect students without a PTSD-diagnosis from the negative feelings and stress they may experience when reading about difficult topics such as sexual assault and racism.

In "Trigger Warnings: From Panic to Data" Francesca Laguardia eloquently summarizes the two main assumptions motivating the adoption of trigger warning. First, there is the assertion that students who have gone through trauma are in need of special protection in order for them to have a fair chance of participation: "The added physiological effects of traumatic triggering can only further imperil these students (or at least their grades), making an already challenging situation impossible for as long as the response lasts" (888). Trigger

warnings are believed to offer the necessary protection by mentally preparing the traumatized students for the content to come and by giving them a chance to avoid it by not doing certain readings or dropping the course. Second, trigger warnings are elevated by their advocates as ways of raising awareness about and normalizing trauma: "These signals offer a lesson to untraumatized students that such students exist, and that such people exist, and that these responses to trauma exist" (890). In brief, the idea is that trigger warnings are a small accommodation to make in order to make the classroom an inclusive and accepting space for those whose experiences may otherwise undermine their learning.

2. The Data

While the intentions behind trigger warnings may be noble, the empirical studies conducted in the past few years on the effects of trigger warnings have brought scientists to a consensus that trigger warnings are not helpful to traumatized and untraumatized students and likely have concerning side-effects for both groups. The most comprehensive data on the effects of trigger warning comes from a Harvard study conducted by Bellet et al. in 2018 on non-traumatised individuals with a pre-registered extension that later replicated its key findings on people who have survived Criterion A trauma as defined by DSM-5. The most charitable conclusion of the 2018 study was that trigger warnings have little effect other than that for some people they cause a slight increase in negative emotions upon receiving the warning: "Trigger warnings do not appear to affect sensitivity to distressing material in general, but may increase immediate anxiety response for a subset of individuals whose beliefs predispose them to such a response" (Bellet et al., 140). The more concerning finding, however, indicates that "trigger warnings may present nuanced threats to selective domains of psychological resilience," namely they may decrease people's confidence in their own mental resilience and facilitate a "disability-related stigma around trauma survivors" (Bellet et al.,

140). While the latter result did not replicate on the 2020 study conducted on trauma survivors (i.e., participants presented with trigger warnings did not proceed to rate other trauma survivors as less mentally resilient than the participants in the control group), the 2020 study confirmed that there is no evidence that trigger warnings are helpful as well as unveiled substantial evidence that trigger warnings reinforce the trauma survivors' view of their trauma as central to their identity (Jones et al., 905). Furthermore, they discovered that those with more severe PTSD symptoms experienced an increase in anxiety after being trigger-warned, remarking that "trigger warnings may be most harmful for the individuals they were designed to protect" (Jones et al., 915). These findings as well as a meta-analysis of other studies led the authors to assess the use of trigger warnings as "irresponsible to victims of trauma" and to caution against it (Jones et al., 915).

Apart from the Harvard scientists' definitive debunking of the supposed benefits to trigger warnings, there have been studies conducted more specifically on students and on more ambiguous rather than explicit content. Such empirics are relevant to consider because college professors have in the past years tended to adorn more and more mild material with trigger warnings guided by the "better safe than sorry" principle. However, trigger warnings are not a harmless measure as shown by a 2021 study confirming the "nocebo hypothesis," namely the idea that trigger warnings lead students to "interpret the upcoming material as more consequential than they would appraise it sans warning" (Bruce et al.). The study measured students' physiological responses to receiving a disclaimer before a relatively mild clip from a Harry Potter movie. When the disclaimer included the phrase "trigger warning," the students tended to display significantly greater physiological symptoms of stress than in the control group leading the researchers to conclude that since "trigger warnings are found to have negligible benefits in addition to anxiogenic effects, there is a reluctance to recommend

their use" (Bruce et al.). A 2019 meta-analysis of 5 experiments in which 1600 participants were exposed to ambiguous stimuli discovers a similarly concerning implication for the salience of employing trigger warnings in the classroom setting. Trigger warnings increased the participants' negative mood, anxiety levels, and negative expectations before they viewed the image and led some to drop out of the study but had "few subsequent benefits" (Bruce et al.). With the data about the lack of beneficial effects and the potential harms of trigger warnings piling up, there is little doubt that instructors ought not implement them in their teaching.

3. The Stakes

Even though the empirical data gives little reason to adopt trigger warnings in the classroom, the fact remains that instructors cannot make the decision solely on the basis of empirics. Students often themselves demand trigger warnings or even the elimination of certain materials. This leaves instructors unsure whether to yield to the students' wishes despite knowing that issuing trigger warnings and permitting avoidance-behaviors is not necessarily benefiting their students. Too often, if they wish to uphold the teaching style that they believe best facilitates a meaningful engagement with course topics, there looms a threat that they will face their student's displeasure that has the potential to negatively affect their careers. A 2017 testimony of an associate professor of Rhetoric at Berkeley Ramona Naddaff "The Wrong Words in the Wrong Times" offers good insight into the practical consequences that enabling the sensitivities of students has on the atmosphere in the classrooms. It shows what exactly is at stake for academia if it chooses to disregard the empirics behind trigger warnings and blindly endorse them for their imaginary symbolic value of signaling that we care and support trauma survivors.

Ramona Naddaff has found herself under criticism of a student whom we will refer to as Tony (pseudonym). In an online reflection written in response to one of the first lectures in her course, Tony expressed the belief that Naddaff's "'condoned' sexual violence and 'reinforced patriarchal structures to further oppress non-men-identifying individuals, including women." Furthermore, Tony felt his instructor "*invalidates, trivializes, and marginalizes* the lived experience of sexual violence survivors" (96). What Tony responded to so strongly, however, is according to Naddaff's testimony, a mere usage of a metaphor common in scholarly discussions of rhetoric, not even central to Naddaff's lecture:

"In passing, I explain the ancient connection between rhetoric and violence, some Greeks associating rhetoric with a violent force that overcomes its listeners' resistance to persuasion. I mention that sometimes rhetoric is metaphorically compared even to rape, a trope of special importance in Gorgias's *Encomium to Helen*. I then show an image of the goddess Peitho (Persuasion), next to which I write "rhetoric as rape," thinking that this formulation, which I have taken from an article, will entice students to consider what it means to compare rhetoric with violence" (93).

Tony's reaction, furthermore, was an outlier rather than a commonly shared take on the professor's lecture. According to another student's testimony, Naddaff's style of teaching was nothing to get upset about: "this was not an act of sexual violence, not a depiction of sexual violence, not pornography, not a story about rape, not even remotely a sexual discussion, but a casual mention of a metaphor, albeit a gritty one, that was relevant to the course" (93). Nevertheless, Tony's accusations achieved a significant shift in the class atmosphere. They prompted the instructors to meet with him individually as well as to conduct an open classroom discussion about the professor's use of language in which some students expressed that "they want trigger alerts" (98). In contrast, other students defended

open discussion through comments such as “I don’t think that we should dismiss generations of scholarship just because there are words we find troubling and uncomfortable. Part of our education here is to be made uncomfortable in our own thoughts and experiences and to learn to develop critical thinking around controversial issues” (98). Nevertheless, the professor’s experience of teaching the class shifts as she reflects: “It takes the class a few weeks to settle down. I can never imagine speaking freely again” (99). Some may view Naddaff’s shift of attitude as too dramatic, but instructors, especially those without the safety of a tenure position, have very valid reasons for being afraid of offending a Tony in their classrooms.

The trigger warning debate is merely a small subset of the larger issue of the growing tendency of US universities to put the comforts of their students above their intellectual development. As Naddaff’s teaching assistant Katharine hints at, a single student like Tony has the capacity to alter the course of his instructors’ careers regardless of whether or not Tony’s response is proportionate to “the crime” committed: “in an academic context in which the “right” not to be emotionally or intellectually “injured” in any way was much discussed and perhaps excessively protected, Tony had actual power” (95). The dilemma that Katharine and Naddaff are navigating is one in which many liberal professors find themselves in. On the one hand, Katharine as a self-proclaimed feminist comprehends “the feeling of self-empowerment that comes with the assertion of personal rights” and “mobilizing for a cause” (99). However, she also worries about “the dangers, in this kind of rights-based thinking, of solipsism and myopia, and worse—a sense of entitlement that has led not to any democratic resolution of the original problem but instead to an identity politics that has become self-centered and personal” (99). She concludes her testimony with a powerful metaphor that captures the dangers of allowing and perhaps even encouraging young people to place one person’s personal emotional comfort on a pedestal and pursue it even when it

negatively affects the educational experience of their peers: “young people are holding their self-protective trigger warnings up like crosses to ward off the devil. To do so risks taking the self too seriously, and the social not seriously enough” (99). Stories like this suggest that classrooms cannot commit to being a “safe space” in the sense that teaching will never evoke strong negative emotions in any one student if classrooms are to remain a “safe space” for discussion, argument, and a daring scholarly exploration of ideas.

B. Campus Climate

1. The Problem of Safetyism

Psychologists and civil rights activists alike have suggested that the rise of popularity of trigger warnings is a mere facet of a larger issue plaguing US academia. Most notably Jonathan Haidt and Greg Lukianoff published an article in the Atlantic in 2015 that due to interest it received became a book-length project. *The Coddling of the American Mind: How Good Intentions and Bad Ideas Are Setting Up a Generation for Failure* explores the idea that phenomena such as trigger warnings, safe spaces, strict campus speech codes, and protesting controversial speakers are on the rise because administrators, educators, as well as students themselves have accepted the proposition that students are mentally fragile and need to be protected not merely from physical but also from emotional distress. Haidt and Lukianoff do not deny that there are real issues such as social inequalities and mental health challenges that students face, issues towards which universities can and should direct resources and attention. However, they worry about what consequences the approach taken by the universities to protect their students has on the mental resilience of the students as well as the future of academia. They emphasize that in psychology it is broadly accepted that "what people choose to *do* in their heads will determine how those real problems affect them" (14) and suggest universities err when they ignore these psychological findings and cater to the students' every demand thus placing their student's comforts over their personal and intellectual growth.

2. The Empirics of Safetyism

Since the publication of *The Coddling of the American Mind*, researchers have begun to empirically study Haidt and Lukianoff's theories. Besides the research that supports their concern about implementing trigger warnings which I presented in the section before, a

recent study from University of California, Irvine establishes the relationship between multiple variables as predicted by Haidt and Lukianoff: "we found that students' self-reported prevalence of cognitive distortions positively predicted their endorsement of safetyism-inspired beliefs, the belief that words can harm, and support for the broad use of trigger warnings" (Celniker et al, 1). This large (N-786), economically and ethnically diverse study finds that the extent to which you engage in cognitive distortion significantly predicts whether you believe that words can harm and whether you support trigger warnings. This leads the authors to suggest that overextending the definition of violence, prejudice, and trauma "may inadvertently engender 'looping effects' whereby students come to interpret actions that they would have otherwise deemed minimally harmful as more aversive" (Celniker et al, 5). Furthermore, the study discovers a negative association between safetyism-inspired beliefs and both resiliency and analytic thinking which the authors find particularly relevant to the dilemmas faced by higher institutions: "if university stakeholders aspire to develop campus cultures and evidence-based policies that better prepare students for the conflict-ridden world they inhabit, then they must be more willing to scrutinize the psychological antecedents of safetyism-inspired beliefs and the consequences of safetyism-inspired practice" (Celniker et al, 5). While more empirical research is needed, especially to establish firm causality, this study gives good reasons to be worried about the pernicious effects of glorifying safety and fostering a culture of oversensitivity.

3. The Pursuit of "Safety" On College Campuses

The idea that speech is violence lies at the very core of the problem US campuses face today. The changes in the way universities approach their students discussed by Haidt and Lukianoff find their ideological roots in accepting intellectual descendants of the proposition that speech is violence such as the idea that you "dehumanize" students by wrongly assuming

their nationality or gender, that you "invalidate" students' identity if you question their experience-based conclusions on social issues, and that you can "harm" students by expressing disagreement with their core beliefs. Once we begin viewing the consequences of distressing and offensive speech as a legitimate threat to students' safety, speech codes aimed at preventing "violent speech" become as important as campus security teams that protect students from physical violence. By discussing some most controversial case-studies of the past years, I hope to show that university communities have accepted the idea that free discussion lies in the way of ensuring a "safe environment" for the students and examine the implications of the shift for the quality of higher education across the US.

Despite the popularity of mockingly portraying liberal students as "snowflakes"—the implication being they are fragile and melt easily—the issue of canceling speech for its purported harmful effects is a non-partisan one. When it comes to attempts at speaker disinvitations, left-leaning crowds have been the predominant source of calls for speech-suppression in the last decade, especially in the tense beginnings of the Trump presidency (see table 1). Nevertheless, the right has quite significantly caught up in the years since and is not trailing far behind (see table 1). As far as critiquing scholars and professors is concerned, the left likewise clearly dominates in the last two years but has been closely followed and even overtook in 2017 by the calls for speech suppression from the right (see table 2). Speaker-disinvitations and scholar controversies started by the right tend to more often be related to a conservative tendency to revere things like marriage, presidency, and pre-natal life and seems to be motivated by a wish to instill similar values onto others or at least not morally corrupt them through exposure to contrary views. Because my main focus is discussing censorship stemming from the idea that some speech is violence and endangers students, I will mostly be bringing up cases where criticism comes from the left. Rest assured

however, that speech censorship from the right exists, is concerning, and ought to be equally decisively rejected.

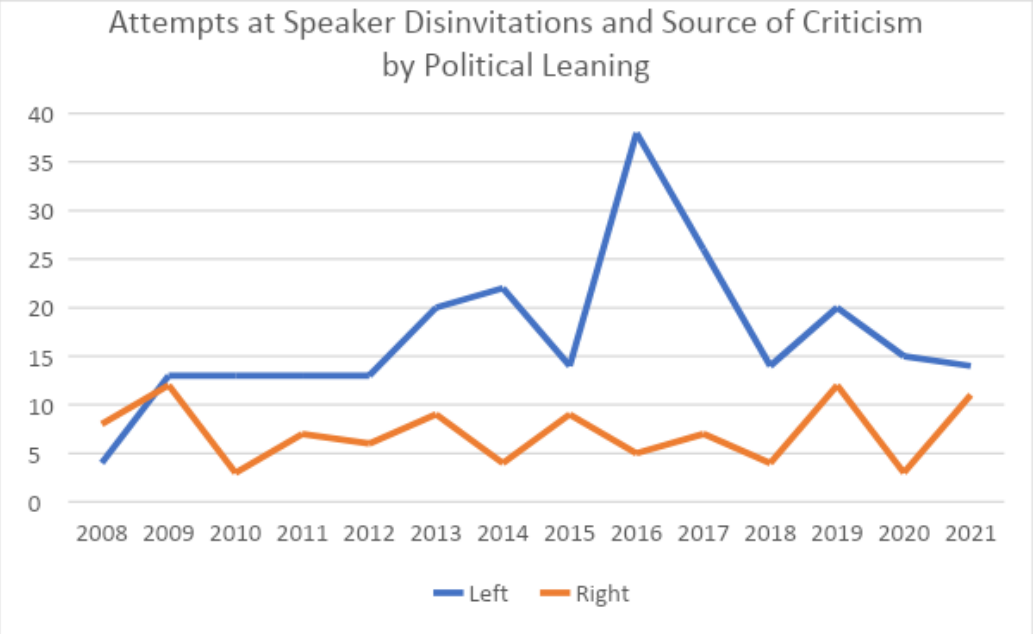

Table 1: Cases of speaker disinvitation incentives on US campuses overtime from the political right and left (data sourced from FIRE's Disinvitation Database)

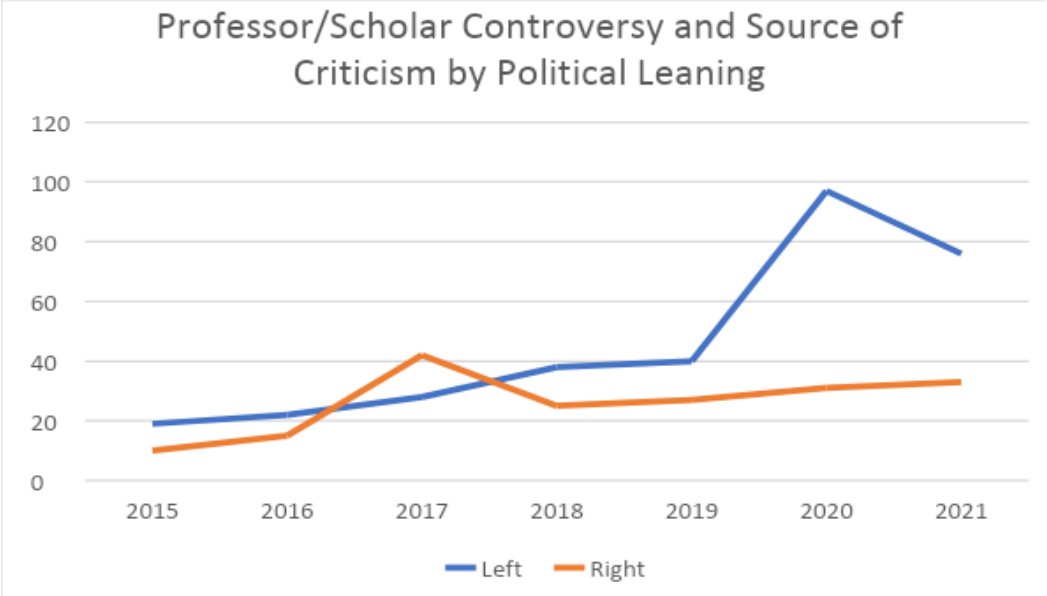

Table 2: Cases of professors and scholars caught into expression-related controversies on US campuses overtime (data sourced from FIRE's Scholars Under Fire Database)

I. "Dangerous" Ideas

Students, educators, and administrators alike have at some point in the past decade made a small and yet monumental shift in discussing racist, transphobic, xenophobic, and other kinds of undesirable ideas. What were once considered ideas that *aim to* dehumanize certain people, *aim to* invalidate their identities, and *aim to* harm the students' *feelings,* are now discussed as ideas that do dehumanize individuals, invalidate their identities, and harm them. Shifting our language ever so slightly fundamentally changes the metaphysical status of humanity and dignity by attributing way too much power to such speech and the morally condemnable individuals spreading it. One's humanity, identity, and personal well-being are things that are constructed and maintained by one's day to day interactions and can by no means be taken away by a couple of demeaning comments from strangers. For example, a group of young men may *aim to* objectify and hurt me by discussing my appearance disrespectfully and shouting sexist remarks and abuses my way, but as long as I *am* treated respectfully by my loved ones on a day-to-day basis, their words will not make me any less of a human being, nor will they make a significant impact on my emotional well-being (assuming I am able to leave their presence when I want to). Members of the university communities, however, have accepted words as much more powerful than they actually are, rendering students, educators, and administrators alike much more likely to react strongly and defensively in face of ideas they view as objectionable.

The most notorious example of what happens when a community embraces the view that words and ideas have power to inflict serious evils onto their members are the violent protests urging for a disinvitation of Milo Yiannopoulos—a famous conservative provocateur and Trump supporter—that happened at Berkeley in 2017. The goal of the protests was to prevent Yiannopoulos's speech from happening and to achieve said goal protesters resorted to

vandalism and violence of extraordinary degrees. The property damage alone amounted to $100,000 on campus and $400,000 to $500,000 off campus (Kutner). However, more concerning even was the violence: multiple people got beaten up, a student journalist trying to record the event got chased off withs sticks and punches (Jandhyala), and a woman was pepper sprayed while giving an interview to a news team (Egelko). The event was indeed canceled, and the aggressive crowd faced little to no consequence for their actions. The police only arrested a single person that night (for failure to disperse) and UC Berkeley refrained from openly disciplining its students. While it is impossible to know how many of the protesters were Antifa activists rather than members of the UC Berkeley's community, some students did come forward and admitted to participating violently yet they faced no consequences.

Most concerning, however, was the way the student body reflected on the protests through UC Berkeley's student led newspaper *The Daily Californian*. A series of five op-eds were published soon after the incidents under the theme "Violence as Self-Defense" with the opinion pieces titled "Violence helped ensure safety of students," "Condemning protesters same as condoning hate speech," and "Black bloc did what the campus should have." The authors of the opinion pieces clearly implicitly or explicitly subscribe to the view that some speech is violence and as a consequence apologize, justify, and downright embrace violence as a valid way of boycotting the Yiannopoulos event. One author proclaims that "Antifa was there to protect UC Berkeley students when the administration was not" (Lawrence) while the other joins in expressing gratitude for the presence of the violent protestors, saying "I'm here to thank the radical measures the AntiFas took to ensure my safety" (Prieto). Behind both of those statements lies a belief that certain kinds of speech because of its content presents a threat to safety that university has a duty to eliminate. If the university fails, the students are

right to take up any means necessary to ensure their own safety. Standing in their way would be equivalent to stifling the students' right to self-defense, as the third op-ed author informs us: "asking people to maintain peaceful dialogue with those who legitimately do not think their lives matter is a violent act" (Dang). The assertion that violence would have been done onto the students if the event was permitted to go on is highly speculative and overblown and yet seems to be unquestioningly accepted by these students. As the fourth opinion writer dramatically asserts, to critique violent protests means to disregard serious damage done unto students by a speaker like Yiannopoulos and shows that "you care more about broken windows than broken bodies" (Meagley). The assumption of danger behind permitting the expression of the speaker's view plays a strong role in their arguments, yet its truth goes uninterrogated.

Berkeley's 2017 protests are perhaps the most extreme example of where we may end up if we commit to viewing controversial speech as violent and harmful. The tensions of the time were certainly exacerbated by the general political atmosphere—only a few weeks before, for example, Trump began his presidency. However, the hesitancy of the university to penalize those causing mayhem as well as the students' endorsement of violence as a valid response to unwelcome ideas are in no way unique or even unusual phenomena on US college campuses. In 2017, a survey by Brookings Institution found that 19% of students agreed that using violence to prevent a speaker from speaking is acceptable (Villasenor). The same year, McLaughlin and Associates found that 30% of students polled believe that violence can be used to stop people from "using hate speech or engaging in racially charged comments" (French). In 2021 a survey done by FIRE and College Pulse corroborated such results by finding that 23% of students from over 150 universities find it rarely (17%), sometimes (5%), and always (1%) acceptable to use violence to stop a campus speech

(College Pulse). Any percentage above 0% should concern us, but to see up to 30% express sympathy with such a radical proposition is staggering.

Unfortunately, administrators have more often than not found it easier to cater to the demands of their "customers" and to endorse the idea that speech harms, despite its implications for the normalization of violence. A good example of how commonplace and ordinary such attitudes towards expression are comes from Colorado State University's campus in early 2022. CSU as a public university has an obligation to uphold student's First Amendment rights, yet its administration found an inventive way to express allegiance with the ideas that speech harms by putting up a sign directing students towards resources at their disposal to cope with free speech: "If you (or someone you know) are affected by a free speech event on campus, here are some resources…" (Soave). A message communicated to the students through such a poster is remarkably similar to that of trigger warnings: you (and those around you) are fragile and need protection from ideas. Such messaging, while well-meaning, is counterproductive to encouraging mental resilience in students as it has the potential to become a self-fulfilling prophecy much like trigger warnings. UC Berkeley's opinion piece writers and CSU's administration assume as a given a highly contestable and politicized interpretation of reality by presenting speech as a hindrance to students' safety and well-being. While the two examples differ in degree, they are of the same kind as they are motivated by the same ideology that presents ideas as literally rather than merely figuratively dangerous.

II. Glorified "Safety"

The implicitly and sometimes explicitly endorsed notion that ideas endanger the safety of students has led universities to foster a culture of hypersensitivity. Haidt and Lukianoff mark what is going on with the term safetyism—the cult of safety—or an

"obsession with eliminating threats (both real and imagined) to the point at which people become unwilling to make reasonable trade-offs demanded by other practical and moral concerns" (32). When it comes to the treatment of speech on college campuses, university communities have indeed become so passionate about creating "safe environments" for their students that they are willing to take extreme actions against any expression perceived by students as harmful without considering relevant factors for rational evaluation of an appropriate response such as the intent of the speaker. Rules of common etiquette that dictate one to interpreting the opponent's views charitably and resolving conflicts respectfully have been frequently discarded as has been a sense of proportionality in the sanctions faced by those whose speech upsets.

One of the biggest controversies that exemplifies oversensitivity is the response ensued from the Yale community after Erika Christakis, a lecturer in developmental psychology and a president of a small residential college, sent an email expressing concern about the significance of Yale's administration warning students not to dress up in insensitive costumes for Halloween in an official email. Christakis's email is a thoughtful exploration of what she admits is a difficult topic to navigate, prompted by "a number of students frustrated by the mass email" (Christakis) who reached out to her and her husband co-presiding the college. She states clearly that her concerns stem from her insights as a developmental psychologist who wishes to allow young adults to exercise their own strength and judgment and not from a "wish to trivialize genuine concerns" (Christakis). Despite her civility and her clearly good intentions, the reaction of a faction of students to her email is markedly negative. Students demanded that the couple be removed from their residential positions and proceeded to attack them "with hateful insults, shouted epithets, and a campaign of public shaming" (Friedersdorf). In an open letter to Christakis, the students took the most

uncharitable if not downright incorrect interpretations of her email claiming it "trivializes the harm done by these tropes [insensitive Halloween attire] and infantilizes the student body to which the request was made" (Friedersdorf) even though Christakis explicitly takes a stance specifically against trivializing and infantilizing. Yale's president responded to the uproar by "acknowledging students' pain and committing to 'take actions that will make us better'" (Haidt and Lukianoff, 56). He did not at all discuss whether the students' responses were proportionate to the offense or more fundamentally whether insults and shouting matches are an appropriate way of conflict-resolution. The university did not express any support for the Christakises until weeks after the event. Eventually, the couple resigned from their positions in the college and Erika from her teaching position as well (Shick). Even if the critique towards administrative paternalism that Christakises voiced was ill-placed, the backlash against them was overblown, uncivil, and disproportionate to their perceived offense.

Many similarly overblown controversies have occurred since 2015 in which students, backed by the administration and faculty, respond disproportionately and unforgivingly to their educators' mistakes. Most often students choose to interpret statements of context, see no value in considering the speaker's intent, and see harsh backlash as not only justifiable but a duty of morally vigilant communities. When in 2020 Rose Salseda, an assistant professor in art history at Stanford, played a clip from the N.W.A song "Fuck the Police" and read some lyrics containing the n-word from her slides, neither the circumstance of her utterance nor her positionality mattered. Despite being a Latinx professor lecturing in a Comparative Studies in Race and Ethnicity who made an arguably ill-considered decision to present a song in the way its authors wrote it, the students were unforgiving. Even though Salseda apologized after being confronted with the students' upset, the Undergraduate Senate harshly condemned her for her "continuous aggressions against the black community" and demanded that all the

professors teaching the course take "cultural humility training" (Thuman). In 2021 students and professors from U Michigan were similarly unforgiving towards professor Bright Sheng and like in Salseda's case considered his attempt to apologize as an even greater reason to reprimand him. Prof. Sheng, an Asian-American, was according to his biography on U Michigan's website a teenager during Chinese Cultural Revolution who narrowly avoided being sent to a re-education farm by auditioning for a music ensemble. The offense that led to him stepping down as a professor was showing a 1965 version of *Othello* which is notoriously controversial because the main character is played by Lawrence Olivier in blackface. Mimicking the rhetoric devices of their students, the composition department's professors condemned Sheng's actions as "disappointing and harmful to individual students in many different ways, and destructive to our community" (Schuessler). Salseda's and Sheng's actions may indeed be seen as inconsiderate to some students, but to depict them as acts of aggression deserving of institutional sanctioning feeds into the ill-conceived idea that ensuring a "safe learning environment" requires an extreme level of sensitivity to any kind of offense or emotional distress.

Perhaps the most representative example of the culture of hypersensitivity fostered by nearly a decade of overblown responses, however, is the firing of an adjunct professor at Fordham in 2021. Christopher Trogan was fired after mixing up the names of two black students in class. He attributed his mistake to his "confused brain" as they were arriving to class late and he claimed to have made attempts to apologize to them: "I have done my best to validate and reassure the offended student that I made a simple, human, error. It has nothing to do with race" (Skelding and Byrne). The plot thickens, however, because it was not merely the students who overreacted, Trogan did too. He wrote a curious email to the students in his class afterwards in which he apologized, stressed "everything he has done for

minorities" as well as encouraged upset students to complain to the administration: "Depending on your response to the officials above, I may—or may not—be your professor in class next week. It's all up to you" (Skelding and Byrne). Given the lack of transparency of the situation it is difficult to identify the greatest victims of the situation—Trogen did get fired, but perhaps his own overreaction contributed to it more than reports can tell. The fact that an innocent name mix-up can start a causal chain that costs an academic his job, however, shows how unforgiving, ruthless, and hostile universities have become in responding to honest mistakes. The culture behind these cases developed overtime as a result of administrators and faculty passively allowing or even actively encouraging a type of conflict-resolution that views its ends of pursuing equality and inclusion as important enough to justify even the most uncivil and unjust means.

III. The True Dangers of Academia

Paradoxically, attempts of universities across the US to facilitate a "safe" learning environment for their students by censoring expression has made universities a dangerous space for intellectual activity of students and professors alike. Too many incidents have occurred over the past decade in which not only careers, but the physical well-being of professors, scholars, and students have come under threat because they expressed controversial ideas or opinions or merely failed to participate in cancel culture.

Activities necessary for the functioning of higher education can put one in serious danger in the current academic environment. In 2017, for example, a violent crowd of students gave Allison Stanger, a professor at Middlebury college, a concussion because she—a self-proclaimed Democrat—was trying to moderate a discussion with Charles Murray who was a scholar at American Enterprise Institute at the time (Stanger). Her intention was to engage Dr. Murray in an intellectual discussion, but because she did not reject Dr. Murray

and his ideas without consideration, she was perceived guilty by association. Kathleen Stock, a philosophy professor at University of Sussex, faced similarly ungrounded threats of violence which in 2021 led her to resign from her position. Like Stanger, Stock herself as a feminist and a researcher into the topics of sexual orientation, gender self-identification, and sexual objectification does not belong to the ideological right. Yet because her research into gender identities was branded by some as transphobic, groups of students were adamant to see her fired at all costs, employing protest, shaming, and intimidation techniques that led her to mark her experience as "medieval" (Adams). More ridiculous still is the suspension that Gordon Klein, a UCLA professor, faced in 2021 because he refused a request to grade students of color differently to other students (Shoaib). If moderating a discussion, researching gender, and maintaining objective grading standards can get you anything from a suspension to a concussion and death threats, universities are not doing enough to ensure a safe working environment for their employees.

Students are in no better position than their educators. Three recent examples demonstrate just how far safetyism stretches. First, there is Austin Tong from Fordham university whose two Instagram posts in 2020 led to him facing disciplinary action because of a purported "bias and/or hate crime and threatening or intimidating behavior" (Patridge-Hicks and Russo). In his first post, Austin shared a photo of a police captain who died while responding to a looting with a caption "Y'all a bunch of hypocrites" while his second offense was sharing a photo of himself holding a gun which he legally owns to commemorate the Tiananmen Square Massacre (Patridge-Hicks and Russo). Regardless of how one evaluates Tong's way of communicating his ideas, it is concerning that universities view their student's activity on their private social media accounts as subject to the university's speech guidelines. The second example comes from University of Virginia's School of Medicine which

sanctioned a student Kiaran Battacharya for raising a question at the panel discussion on the subject of microaggressions and consequently "branded him a threat to the university and banned him from campus" (Soave). The question that prompted the chain of events leading to Battacharya's suspension was a clarificatory question on whether one must be a member of a marginalized groups to be a victim of a microaggression. Similar to his case, Adrianna San Marco from Syracuse University faced disproportionately grave consequences because she critically engaged with a controversial topic. San Marco was fired from her columnist position at a student newspaper *The Daily Orange* because she wrote an opinion piece dismissing the notion of "institutional racism" for an unrelated website (Flood). She received threats and harassment from her fellow students as well as physical threats promising violence upon her return to campus and has had peers and professors celebrate her firing as well as call for her expulsion (Flood). The overwhelming message of such examples is one of warning for students to tread lightly because a social media post, a pointed question, or a critical opinion piece can put them in danger. The effects cannot but be chilling.

The examples I have so far outlined may lead one to believe that students, faculty, and administrators on the ideological left are always the ones to appeal to safety and call for censorship on the basis of intellectually flawed (or at least highly questionable) assumptions such as that expressing some ideas is an aggression and an act of violence. While the ideology of speech = violence has strong roots in postmodernist conceptions about language that made its greater impression on the political left, I want to push back against characterizing this as a simple partisan issue. It has become more and more frequent for those on the political right to present censorship of expression as a means of protecting students from dangerous ideas, though there are often other values than merely safety at play.

For example, in 2017, Essex County College first suspended and then fired its professor Lisa Durden for her appearance on Fox News in which she defended the right of groups such as BLM to organize black-only events (Schmidt). The college justified its decision by emphasizing that they reject "any conduct that implies that all students are not welcome to participate in, or benefit from, our programs or activities on the basis of their race, color, orientation, or national origin" (Schmidt). While Durden's opinions and the way she expressed them lend themselves to much critique, the implication that her conduct was a threat to students because it suggested they do not have a right to equal educational opportunities is a rather farfetched, certainly uncharitable, and a relatively catastrophized interpretation of her speech. Other teachings of Critical Race Theory (CRT) similar to Durden's thoughts on self-segregation have in the past five years increasingly frequently faced similar critiques of being too harmful to deserve a platform. Most passionate uproar regarding CRT arose from controversial ways in which ideas of race and privilege have been passed onto minors and rightfully so—governments have an obligation to ensure certain protections to underage persons, including by determining what does and does not belong on a public school's curriculum. Lately, however, it has been suggested that teaching CRT ought to be banned from higher education as well. Most notably, lt. gov. of Texas recently proposed to remove tenure in publicly funded higher education facilities to enable sanctions to those who teach CRT (NBCDFW). Patrick justifies his decision in a Tweet by saying that professors "poison the minds of young students" with CRT, thus exemplifying that those with right-wing or conservative views are not immune to attributing expression too much power and adopting protectionist and censorial attitudes towards expression when convenient.

Playing into the misconception that some or other kinds of expression can enact violence, harm, and poison minds has a non-partisan allure as it carries with it a ready-made

conclusion that it is permissible and desirable to censor it. Allowing adults—especially young ones—to be exposed to all kinds of ideas and take responsibility for the beliefs they adopt is much more difficult because it requires us to trust people's ability to intellectually self-govern. As foolish and optimistic as such trust may appear at the surface level, however, higher education cannot fulfill its role without it.

4. A Beacon of Hope: University of Chicago Principles

In response to a series of incidents at which students called for disinvitations of controversial speakers, University of Chicago formed a Committee on Freedom of Expression in 2014 that formulated and published a statement intended to offer robust protection of freedom of expression at the university and to acknowledge the centrality of the commitments to free expression for the goals of the university. As of December 2021, FIRE reports on 83 institutions that have joined in endorsing the Chicago Statement (FIRE). Besides voicing the university's commitment to debate, the statement puts emphasis on the duty of universities to create an environment prioritizing intellectual pursuits over comfort: "education should not be intended to make people comfortable, it is meant to make them think. Universities should be expected to provide the conditions within which hard thought, and therefore strong disagreement, independent judgment, and the questioning of stubborn assumptions, can flourish in an environment of the greatest freedom" (University of Chicago, 1). There are many competing ideas about the purpose of higher education. Some insist that the central goal of universities ought to be the pursuit of truth while others discuss the telos as sparking curiosity or fostering critical inquiry. All of those values, however, are centered around learning that is necessarily infringed on if members of university communities demand that certain ideas and topics, however controversial, be excluded from the spectrum of permissible.

The Chicago Principles oppose content-based regulation that ends discussion: "debate or deliberation may not be suppressed because the ideas put forth are thought by some or even by most members of the University community to be offensive, unwise, immoral, or wrong-headed" (University of Chicago, 2). Moreover, they showcase precisely the kind of faith in the ability of people to intellectually self-govern that those looking to stifle speech for protectionist purposes deny: "It is for the individual members of the University community, not for the University as an institution, to make those judgments for themselves, and to act on those judgments not by seeking to suppress speech, but by openly and vigorously contesting the ideas that they oppose" (University of Chicago, 2). As such the Chicago Statement offers a good example of the attitude universities should adopt when addressing speech incidents in order to foster a flourishing intellectual environment where students are treated as capable and responsible thinkers rather than as emotionally fragile victims.

The foresight of adopting such a statement in 2014 is admirable, but there are areas in which the Chicago Statement could say more. While it takes a general stance against content-based suppression of ideas, it does not say anything specific about its commitments to support their faculty in cases when non-malicious word-choices or teaching techniques spur an uproar amongst the students. The numerous incidents such as I outline in earlier sections often make professors, especially the ones teaching controversial subjects or topics, hesitant or even afraid of performing their duties because any perceived mistake, however small, can end up on their records and set them back when seeking a permanent position or a promotion. Even if an improved statement was to mention their commitment to protecting instructors from unjust accusations, however, the fact remains that such a statement is a mere aspirational one, which means it does not bind institutions to translate it into practical policy changes. The systems of dealing with complaints at universities are obscure and designed to

make as little information as possible publicly accessible. Unless controversies gain national publicity, it is not at all difficult to showcase high flying principles while in practice continuing the "customer is king" strategy of addressing speech issues that uncritically caters to students' demands.

C. "Violent" Views and Freedom of Academic Inquiry

A grander and a much older tradition equivalent to the desire to pursue 'safety' by stifling expression on campuses and in the classroom are tendencies of activist movements to demand for censorship when they perceive certain scientific findings as harmful to the groups on whose behalf they advocate. Just as student protestors shouting down speakers rarely take the time to understand the speaker's views, such activists often fail to accurately represent the scientists they critique. Along the same lines, both groups show utter disregard for the speaker's or the scientist's intent. Ironically, the motivations of both students and the majority of activists who join the bandwagon of condemnation are most often good: while some stoop to purposeful misrepresentation or smear-campaigns, the majority merely wish to protect groups of individuals that painful history has made well-deserving of protection. Yet by declaring certain lines of inquiry as off-limits, such an anti-scientific way of pursuing social justice endangers the freedom of academic inquiry and in the long term does more harm than good for the groups it strives to protect. Alice Dreger, a former professor of bioethics at Northwestern University and a life-long activist herself, explores cases in which activism endangers the freedom of academic inquiry in *Galileo's Middle Finger*. With the help of Dreger's work, I will delve into three case-studies of the controversies regarding E.O. Wilson, Napoleon Chagnon, and Michael Bailey in order to highlight the way misguided protectionism affects the state of academia.

Importantly, however, condemning some kinds of protests and some forms of activism must never be taken as a blanket dismissal of protest, social justice, and activism. Dreger certainly subscribes to the idea that it is important for the public to carefully monitor science and voice concerns where concerns are due. In fact, she insightfully presents science and social justice efforts as necessarily co-dependent: "Science and social justice require each

other to be healthy, and both are critically important to human freedom. Without a just system, you cannot be free to do science, including science designed to better understand human identity; without science, and especially scientific understandings of human behaviors, you cannot know how to create a sustainably just system" (17). Activism needs to recognize the importance of scientific inquiry into identity and human behaviors; only solutions informed by an understanding of intra-group dynamics and identity can in the long term better the treatment of indigenous communities, people of color, and transgender individuals. The examples that will follow are cases in which few individuals hastily wishing to discredit a successful scientific figure manipulate and weaponize people's inclinations to stand against injustice and protect a vulnerable minority group from harm. They carry a warning against joining a witch hunt upon insufficient evidence and demonstrate how easily pursuing 'safety' can be used to stifle academic inquiry in areas where science is most needed to separate the wheat from the chaff and lead us to a better understanding of ourselves and those we most differ from.

1. E.O. Wilson

Edward O. Wilson (1929-2021) was a renowned biologist and an expert at the forefront of many groundbreaking discoveries on biodiversity, insects, and human nature. The controversial years in his career, as Carl Zimmer writes in his obituary, began when Wilson took the vast knowledge of evolutionary principles which he accrued through studying insects and applied it to the investigation of other animals and subsequently people. In 1975, Wilson published *Sociobiology: The New Synthesis* which established him as the father of this new field. While Wilson was careful to warn that applying evolutionary principles to humans would not be simple and even called for a discipline of "anthropological genetics," some dogmatically condemned sociobiology and with it Wilson and marked it as an advocacy for

biological determinism that was historically used to excuse oppression and even genocide (Zimmer). Never mind that Wilson's work was a professional, nuanced, and a careful foray into thinking about humans as evolved beings and that he made great efforts to emphasize that sociobiology offers no excuse for racism, sexism, and antisemitism. Mere guilt by association was enough to brand him as an enemy.

Those most convinced by the unprofessional and poorly-reasoned smear-campaign against Wilson's work not only denounced the work but took pains to harass the scientist. In 1978 members of self-proclaimed "International Committee Against Racism" rushed on the stage when Wilson was about to speak and chanted "Racist Wilson, you can't hide, we charge you with genocide!" (Zimmer). Dreger reports that Wilson having recently broken his leg "was in a cast that stretched almost from his ankle to his hip" (139) which the protesters took little notice to as they proceeded to dump ice water on him shouting "Wilson you are all Wet!" (Zimmer; Dreger, 139). Even his Harvard colleagues, most notably Richard Lewontin and Gould, publicly accused Wilson of "promoting a dangerous right-wing science" (209). Rather than actually engaging with Wilson's writing, such colleagues misconstrued implications of Wilson's scientific work and in the classic example of a straw man fallacy dogmatically condemned Wilson for the crimes he has not committed. Under the pretense of protecting groups historically affected by eugenics and determinist thinking, such a public smear campaign wrought havoc in the life of a respectable scientist. As he reflects in an interview with Dreger: "I couldn't sit by and let them say something that was in fact declaring me a racist and a proto-Nazi. I couldn't say, "No comment." I just wasted enormous amounts of energy and just pure time I could have used for something much more valuable" (211). E.O Wilson regained some level of public trust once the scientists who took time to look into his ideas began researching the ways genes influence behaviors in human and

non-human species. But the development of the field of sociobiology has certainly been slower because it began in unnecessary controversy.

2. Napoleon Chagnon

According to Dreger the controversy regarding Napoleon Chagnon demonstrates what happens when truth is not merely overlooked but crushed for political purposes. She calls this disregard for evidence "dangerous intellectual rot occurring within certain branches of academe" and wittily describes Chagnon's case as a situation in which "liberal hearts bleed so much that brains stop getting enough oxygen" (129). Indeed, the story is bizarre and reads like an adventure melodrama: a scientist investing decades of his life into the study of the Yanomamö tribe living in the jungle between Brazil and Venezuela retrospectively receives false accusations of harming the indigenous people through his work and his life and career gets completely upheaved because organizations that should stand behind protocol-abiding scientists such as the American Anthropological Association (AAA) are too eager to present themselves as protectors of the indigenous communities to fact-check the accusations before condemning and ostracizing a fellow academic. For the purposes of this work, I will not delve into all the exciting details of the situation that clear Chagnon's name and explain how (and why) Patrick Tierney, the author of the libelous book *Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon (2000)*, distorted reality and accused Chagnon of purposefully spurring wars, withholding medical care, and causing a measles outbreak among the Yanomamö people just to study them. The details are well recorded in Dreger's *Galileo's Middle Finger* and elsewhere. Most relevant to the question of freedom of academic inquiry is the response of organizations such as AAA to the accusations as it demonstrates how good intentions of protecting a vulnerable group can be destructive to the

field that does most to progress the society's understanding of indigenous people, their ways of life, and the value of protecting their interests.

After the publication of Tierney's book, the AAA called together a taskforce to look into the case and passed a quick judgment condemning their colleague and his work at a conference without giving his side any proper consideration: "although Chagnon was obviously being put on trial at the AAA, no one from the association ever issued him a formal invitation to defend himself. He was to be tried in absentia" (132). Jane Hill, the chair of the taskforce, in response to a fellow scientist who expressed concern about AAA's hasty condemnation of Chagnon elaborates on the reasons for such a rapid decision:

"The book is just a piece of sleaze, that's all there is to it (some cosmetic language will be used in the report but we *all* agree on that). But I think the AAA had to do something because I really think that the future of work by anthropologists with indigenous peoples in Latin America—with a high potential to do good—was put seriously at risk by its accusations, and silence on the part of the AAA would have been interpreted as either assent or cowardice. Whether we're doing the right thing will have to be judged by posterity" (Dreger, 162).

Hill's message demonstrates that the decision of the AAA was well-intended, but also that it was clearly influenced by the external pressures placed onto them by a kind of activism that believes you must instantly decide to be either for the cause or the cause's enemy. Rather than allowing people the time to examine the work of the accused and decide whether or not condemnation is in order, individuals, college administrations, and larger organizations face extreme pressures to immediately take a stance. When the line between science and politics blurs too much, a task force created to *examine* the accusations absurdly feels it must condemn the accused and do it immediately. By condemning a scientist upon insufficient

evidence and upon knowing the main source of accusations against him to be "a piece of sleaze," an organization such as the AAA trades in the values of science and inquiry for virtue signaling. Similarly, in the college setting, institutions of learning feel pressed to condemn instructors the moment they face an accusation of hurting a student's feelings. Protecting freedom of speech in higher education does not mean administration must stand behind just about any conduct of their instructors. Likewise, upkeeping the values of science does not require the likes of AAA and Chagnon to entirely disregard social concerns while pursuing the truth. As Dreger reports, Chagnon himself sought a balance between free inquiry but also socially aware inquiry: "when he found out that the data he had collected on Yanomamö infanticide might be used by the Venezuelan government against them, he had essentially withdrawn the data /…/ this was a scientist out primarily for truth, but never at the cost of justice" (138). Examining and working to prevent ways in which science can be used to propagate injustice bears incredible importance, but condemnations of scientific endeavors ought to be grounded in strong evidence and carefully considered. If those standing up for indigenous people are not guided by a care for truth, evidence, and critical thinking, their protectionism is either a mere performance or an ugly exercise of power.

3. Michael Bailey

In the case of Michael Bailey, a psychologist at Northwestern University, the activists feigning to protect the transgender community by starting a crusade in fact appeared to be motivated by the desire to have absolute power and control over the truth about transgender identities. In Dreger's words, they "tried to bury a politically challenging scientific theory by killing the messenger" and in the process "charged Bailey with a whole host of serious crimes, including abusing the rights of subjects, having sex with a transsexual research subject, and making up data" in order to discredit him enough to bury his ideas (15). The

undesirable ideas Bailey expressed in his book *The Man Who Would Be Queen (2003)* were from the current perspective on identity and sexuality extraordinarily progressive. He opposed the traditional rather heteronormative understanding of transgenderism as a mere glitch in the binary in which women end up trapped in male bodies and men trapped in female bodies (Dreger, 15). By following the data he accumulated through years of working with and interviewing transgender people, Bailey suggested that "in cases of men who become women, transgender isn't just about gender identity, but also about sexual orientation—about eroticism" (Dreger, 14). Bailey emphasized that by giving greater nuance to the picture of trans identity, his work advocates for the greater acceptance and a normalization of transgender individuals and their wishes: in his book he "unequivocally supports the right of all people to be gender-variant, to enjoy whatever sexual orientations they have (so long as they're not using anyone who can't consent or hasn't consented), to be recognized by the gender labels they choose for themselves, and to get whatever medical interventions they wish" (Dreger, 82). Partly because most of those condemning him did not read his work, however, his attempts to forward the transgender cause were condemned as transphobic.

To those interested in the full details of how Bailey's main opponents bent the truth to manufacture outrageous and entirely false allegations about him I yet again suggest reading Dreger's book. For my purposes, Bailey's story demonstrates that those willing to manipulate people's inclinations to passionately stand up for injustice wield a dangerous amount of power in our fast-paced and informationally chaotic society. Bailey's case renders it explicit how the propagators of false accusations exploit a popular heuristic to trust the victims and not fact-check every controversy as it arises. Such a heuristic stems from a very understandable calculation: failing to protect the victims because of a hesitancy to believe

them often in the long run produces more harm than believing instantly and correcting yourself later on the occasion of a false positive. When both the stigma on challenging scientific, social, or political authorities and the risk that victims undertook by voicing their pain were much greater than today, such a heuristic might have worked splendidly. As the cases of E.O. Wilson, Chagnon, and Bailey as well as many similar controversies that have occurred since demonstrate, however, believing instantly and immediately and loudly supporting condemnation whenever a vulnerable group's safety is in question no longer presents a responsible course of action from an epistemic standpoint.

Moreover, contrary to what the most zealous activists might wish us to believe, immediately condemning the accused without looking into the details of the situation and examining the evidence is neither the only nor the best course of action to take for someone genuinely invested in social justice. To demonstrate this point imagine you get into a dispute with your tax-filing service—they claim you owe more tax than you believe you do. You message two friends letting them know you are besides yourself because a tax service wants you to pay too much tax. The first friend replies: "I'm 100% on your side in this. They are absolutely horrible. I'm posting bad reviews on their website and social media as we speak." The second friend replies: "I'm sorry you feel that way. What precisely is the disagreement about? Send me the details and I'll look into it—it could just be a misunderstanding." Which one of the two friends strikes you as more genuine in their care about you? While the first friend enthusiastically and immediately voices their support for you, he does not seem to care to invest the effort into discovering the truth about the situation. Some might say that the only truth that should matter to him is that you are his friend and therefore worthy of unconditional support, but if you happen to be mistaken in your judgment and your friend does not care to learn more, his support might encourage you to commit tax fraud. The second friend, I argue,

responds better because he expresses sympathy for your feelings but also a care for learning the relevant facts that will allow him to take proper action in supporting you once he knows that pursuing the dispute is within your long-term best interests. In the world of social justice activism, especially the one concerned with watching over science, a sincere concern for truth is of the utmost importance for protecting the interests of the vulnerable groups in the long-term.

Conclusion

After looking at the history of free expression, weighing philosophical arguments for and against its regulation, examining insights from evolutionary psychology and neuroscience, and heeding the warnings presented by censorial attitudes of youth and academia, I have arrived at the following conclusion: by far greater dangers lie in censoring expression then in permitting expression to be almost absolutely free. The key foundation of my reasoning and, in my opinion, the most important front on which the battle for free expression will be fought in this age, is the notion that language is not violence, words do not wound, and expression does not do harm.

What does it mean to reject the "speech=violence" fallacy? Does it require you to reject the idea that words are powerful tools of coercion? That expression can lead to negative consequences? That verbal abuse and name-calling are not serious issues? Not in the slightest. I do not deny the significant role language plays in exercising power over people, in organizing crime, and in creating psychological distress, to name only a few examples. I do insist, however, that between hate speech and a hate crime there is a decision to act. Choice stands in between propaganda and genocide and separates reading online forums from act of school shooting. The pen is indeed mightier than the sword: compelling ideas can convince many to act simultaneously while brute force coerces only the few within its reach. But compared to the sword, the pen's success rate is low: it only moves those already predisposed to choose the action it advocates for. Harms inflicted by expression are almost never directly caused and inevitable, but mediated by a decision, a choice. By restricting freedom of expression, we thus rob people of the liberty to choose what to do. Letting people decide what to do is highly dangerous, but it is also what grounds all other human liberties.

Many worry that claiming that speech cannot be violence because it is mediated by choice implies that when words cause us psychological pain and distress, we choose said hurt and inflict it onto ourselves. It would seem that rejecting the "speech=violence" fallacy requires us to tell a victim of degrading racial slurs, a survivor of sexual abuse experiencing intense anxiety when she reads a description of a rape case in a textbook, or even someone suffering from repeated verbal abuse to just "suck it up," "toughen up," and stop overreacting to "mere words." While some have indeed drawn such connections, I reject the notion that such are the necessary implications of rejecting the "speech=violence" fallacy. There are cases in which the time to make a choice and form a decision before acting is so restricted that the onus lies on the initiator of an action to foresee a response. For example, the man who throws the first punch is responsible for spurring actions of self-defense in its opponent. While the opponent could have in theory chosen not to retaliate violently, we do not blame him for responding with a punch. Like a person defending themselves, victims of "just words" are not necessarily to blame for the pain they suffer as their emotional reactions are often as natural as self-defense.

Adopting the notion that choice mediates the relationship between speech and violence does not require us to "blame the victim" and guilt people whenever they experience distress, offense, and pain as a result of someone's words. What it allows us, however, is to offer them an empowering reminder that they are capable of affecting the way they process the words intended to harm them. You cannot give people the mental tools to make a bullet-wound disappear or ways of thinking that will take away the pain of a nasty stab wound and make it heal without any scarring. By training mental resilience, however, you can give people the knowledge and skills with which they can greatly lessen the pain that expression brings about. Even more importantly, however, by changing their understanding

of language's role in our lives, people can in the long term prevent traumatic experiences from negatively affecting their quality of life. Even the most nasty, despicable, and morally abhorrent words are just words; verbal virulence, no matter how intense, does not have the power to objectively undo people's worth and dignity.

It is crucial, therefore, that we reject the notion that speech equals violence, but that we do it without denying empathy and understanding to those who are negatively affected by expression. Especially when it comes to issues close to people's hearts such as racial justice, issues of gender identity, sexual orientation, and religious beliefs, it is important that we find a way to discuss words as "just words" without underscoring or misunderstanding the emotional pain speech can bring about. Defending free expression not only defends the rights of the offended, the hurt, and the upset to argue back, but also grounds the right of bystanders to openly debate, question, and point out the flaws of distressing views. Universities in particular have the duty to allow for the discussion of even unwelcome, offensive, and disagreeable ideas and to teach their students that they have a choice how to interact with ideas that may evoke pain in them. If young people are not trained in mental resilience and encouraged to adopt the saying "sticks and stones may break my bones, but words will never hurt me," it is unsurprising that they see it as justifiable to throw sticks and stones at those whose words they disagree with.

It is very easy to look at freedom of expression and see in it an obstacle to a harmonious society, an excuse to be rude, uncivil, and to spread hatred. It is just as easy to forget that free expression topples tyrannical governments, exposes and rights injustices, propels scientific development, inspires artistic innovation, and grounds our very ability to hope, dream, and self-determine. The concern over the appropriateness of what is expressed too often detracts our attention from the value we lose when ideas are suppressed. Censorship

and regulations of expression demand too high a price for a kind of safety that we should not desire as it becomes attainable only when we eliminate human freedom.

Works Cited

Adams, Richard. "Kathleen Stock says she quit university post over 'medieval' ostracism." *The Guardian*. 3rd Nov. 2021, https://www.theguardian.com/education/2021/nov/03/kathleen-stock-says-she-quit-university-post-over-medieval-ostracism, Accessed 12th March 2022.

ARTICLE 19. *Responding to "Hate Speech": Comparative Overview of Six EU Countries*. ARTICLE 19, 2018, https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf.

Balot, Ryan K. "Free Speech, Courage, and Democratic Deliberation." *Free Speech in Classical Antiquity*, edited by Ineke Sluiter and Ralph M. Rosen, Penn-Leiden Colloquium on Ancient Values, 2004.

Bartlett, Larissa, et al. "A Systematic Review and Meta-Analysis of Workplace Mindfulness Training Randomized Controlled Trials." *Journal of Occupational Health Psychology*, vol. 24, no. 1, 2019, pp. 108–126., doi:10.1037/ocp0000146.

Bellet, Benjamin W, et al. "Trigger Warning: Empirical Evidence Ahead." *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 61, 2018, pp. 134–141., doi:10.1016/j.jbtep.2018.07.002.

Booth, Joanne et al. "Evidence of Perceived Psychosocial Stress As a Risk Factor for Stroke in Adults: A Meta-Analysis." *Bmc Neurology*, vol. 15, no. 1, 2015, pp. 233–233. doi:10.1186/s12883-015-0456-4. Accessed 12 Feb. 2022.

Breuer, Josef, and Sigmund Freud. *Ueber Den Psychischen Mechanismus Hysterischer Phänomene : (vorläufige Mittheilung)*. Veit, 1893.

Bridgland, Victoria M E, et al. “Expecting the Worst: Investigating the Effects of Trigger Warnings on Reactions to Ambiguously Themed Photos.” *Journal of Experimental Psychology. Applied*, vol. 25, no. 4, 2019, pp. 602–617., doi:10.1037/xap0000215.

Bruce, Madeline J, et al. “Students’ Psychophysiological Reactivity to Trigger Warnings.” *Current Psychology*, (20210527), 2021, doi:10.1007/s12144-021-01895-1.

Cameron, Daryl C., et al. “The Emotional Cost of Humanity: Anticipated Exhaustion Motivates Dehumanization of Stigmatized Targets.” *Social Psychological and Personality Science*, vol. 7, no. 2, 2016, pp. 105–12.

Carter, D. M. “Citizen Attribute, Negative Right: A Conceptual Difference Between Ancient and Modern Ideas of Freedom of Speech.” *Free Speech in Classical Antiquity*, edited by Ineke Sluiter and Ralph M. Rosen, Penn-Leiden Colloquium on Ancient Values, 2004.

Celniker, Jared B, et al. “Correlates of ‘Coddling’: Cognitive Distortions Predict Safetyism-Inspired Beliefs, Belief That Words Can Harm, and Trigger Warning Endorsement in College Students.” *Personality and Individual Differences*, vol. 185, 2022, doi:10.1016/j.paid.2021.111243.

Christakis, Erika. “Email From Erika Christakis: ‘Dressing Yourselves.’ Email to Silliman College (Yale) Students on Halloween Costumes” *FIRE*. 30th Oct. 2022, https://www.thefire.org/email-from-erika-christakis-dressing-yourselves-email-to-silliman-college-yale-students-on-halloween-costumes/, Accessed 12th of March 2022.

College Pulse/FIRE. “2021 College Free Speech Rankings Data.” *Tableau Public*. 24th Aug. 2021, https://public.tableau.com/app/profile/college.pulse/viz/2021CollegeFreeSpeechRankingsData/2021CollegeFreeSpeechRankingsData, Accessed 9th of March 2022.

Dang, Nisa. “Check your privilege when speaking of protests.” *The Daily Californian.* 7 Feb. 2017, https://www.dailycal.org/2017/02/07/check-privilege-speaking-protests/, Accessed 9th of March 2022.

Dawkins, R. (1983) Universal Darwinism. In Evolution from Molecules to Men (ed. D. S. Bendall). Cambridge: Cambridge University Press. pp. 403–25.

DePauw University. Draft of a “Statement of DePauw University Values on Freedom of Expression.” Email to the author, 8. Mar. 2022, provided by Dr. Sarah Steinkamp.

Dreger, Alice. *Galileo's Middle Finger : Heretics, Activists, and the Search for Justice in Science*. E-Book ed., Penguin Press, 2015.

Dunn, John. *The Political Thought of John Locke: An Historical Account of the Argument of the “Two Treatises of Government.”* Cambridge University Press, 1969.

Egelko, Bob. “Woman pepper-sprayed at 2017 Milo Yiannopoulos protest in Berkeley may sue, court rules.” *San Francisco Chronicle*. 13 Jul. 2020, https://www.sfchronicle.com/bayarea/article/Woman-pepper-sprayed-at-2017-Milo-Yiannopoulos-15405392.php, Accessed 9th of March 2022.

Feinberg, Joel. *Offense to Others*. Oxford University Press, 1985.

Feldman Barrett, Lisa. “When Is Speech Violence?” *The New York Times*, 14 July 2017, https://www.nytimes.com/2017/07/14/opinion/sunday/when-is-speech-violence.html, Accessed 12 February 2022.

FIRE. “Chicago Statement: University and Faculty Body Support.” *FIRE*. 2nd Dec. 2021, https://www.thefire.org/chicago-statement-university-and-faculty-body-support/, Accessed 12th March 2022.

Flood, Brian. “Student journalist fired for calling institutional racism a ‘myth’ speaks out: ‘I stand by my analysis.’” *Fox News*. 18th Jun. 2020,

https://www.foxnews.com/media/syracuse-student-journalist-fired-institutional-racism-myth, Accessed 12th March 2022.

French. David. "NRO: NEW COLLEGE STUDENT SURVEY: YES, SPEECH CAN BE VIOLENCE." *McLaughlin & Associates.* 11th Oct. 2017, https://mclaughlinonline.com/2017/10/11/nro-new-college-student-survey-yes-speech-can-be-violence/, Accessed 9th of March 2022.

Friedersdorf, Conor. "The New Intolerance of Student Activism." *The Atlantic*. 9th Nov. 2015, https://www.theatlantic.com/politics/archive/2015/11/the-new-intolerance-of-student-activism-at-yale/414810/, Accessed 12th March 2022.

Friedman, Richard A. "The Neuroscience of Hate Speech." *The New York Times*, 31 Oct. 2018, https://www.nytimes.com/2018/10/31/opinion/caravan-hate-speech-bowers-sayoc.html.

Gibson, Tobias T. "Bad Tendency Test." *The First Amendment Encyclopedia*, 2009, https://www.mtsu.edu/first-amendment/article/893/bad-tendency-test.

Haidt, Jonathan, and Greg Lukianoff. *The Coddling of the American Mind*. Allen Lane, 2018.

Hansen, Mogens Herman. *The Athenian Democracy in the Age of Demosthenes: Structure, Principles and Ideology*. Translated by J.A. Crook, Blackwell, 1991.

Harris, Lasana T., and Susan T. Fiske. "Dehumanized Perception: A Psychological Means to Facilitate Atrocities, Torture, and Genocide?" *Zeitschrift Für Psychologie*, vol. 219, no. 3, Jan. 2011, pp. 175–81.

Henrich, Joseph. "Selective Cultural Processes Generate Adaptive Heuristics." *Science*, vol. 376, no. 6588, 2022, pp. 31–32., doi:10.1126/science.abo0713.

Heller, Brittan, and Joris van Hoboken. *Freedom of Expression: A Comparative Summary of United States and European Law*. Transatlantic Working Group, 3 Mar. 2019.

Herodotus, and A. D. Godley. *The Histories*. Harvard University Press, 1920.

Hoffman, Morris B. "Neuroscience Cannot Answer These Questions: A Response to G. and R. Murrow's Essay Hypothesizing a Link between Dehumanization, Human Rights Abuses and Public Policy." *Journal of Law and the Biosciences*, vol. 3, no. 1, Apr. 2016, pp. 167–73, https://doi.org/10.1093/jlb/lsv041.

Jandhyala, Pranav. "Anarchists at the Berkeley riot punched me in the face and tried to steal my phone." *The Tab*. 2017, https://thetab.com/us/uc-berkeley/2017/02/02/anarchists-milo-berkeley-riot-3271, Accessed 9th of March 2022.

Jiang, Yanping, et al. "Basal Cortisol, Cortisol Reactivity, and Telomere Length: A Systematic Review and Meta-Analysis." *Psychoneuroendocrinology*, vol. 103, 2019, pp. 163–172.

Jones, Payton J, et al. "Helping or Harming? The Effect of Trigger Warnings on Individuals with Trauma Histories." *Clinical Psychological Science*, (20200601), 2020, doi:10.1177/2167702620921341.

Kutner, Max. "Inside the Black Bloc Protest Strategy That Shut Down Berkeley." *Newsweek.* 24 Feb. 2017, https://www.newsweek.com/2017/02/24/berkeley-protest-milo-yiannopoulos-black-bloc-556264.html, Accessed 9th of March 2022.

Laguardia, Francesca, et al. "Trigger Warnings: From Panic to Data." *Journal of Legal Education*, vol. 66, no. 4, 2017, pp. 882–903.

Lawrence, Neil. "Black bloc did what campus should have." *The Daily Californian*. 7 Feb. 2017, https://www.dailycal.org/2017/02/07/black-bloc-campus/, Accessed 9th of March 2022.

Liebsch, Burkhard. "What Does (Not) 'Count' As Violence: On the State of Recent Debates About the Inner Connection between Language and Violence." *Human Studies*, vol. 36, no. 1, 2013, pp. 7–24.

Locke, John. *Second Treatise of Government and A Letter Concerning Toleration*. Edited by Mark Goldie, Oxford University Press, 2016.

McGinnis, John O. "The Once and Future Property-Based Vision of the First Amendment." *The University of Chicago Law Review*, vol. 63, no. 1, 1996, pp. 49–132.

Meagley, Desmond. "Condemning protesters same as condoning hate speech." *The Daily Californian.* 7 Feb. 2017, https://www.dailycal.org/2017/02/07/condemning-protesters-condoning-hate-speech/, Accessed 9th of March 2022.

Mill, John Stuart. *On Liberty*. Edited by David Bromwich and George Kateb, Yale University Press, 2003.

Mchangama, Jacob. *Free Speech : A History from Socrates to Social Media*. First ed., Basic Books, Hachette Book Group, 2022.

Momigliano, Arnaldo. "Freedom of Speech in Antiquity." *Dictionary in the History of Ideas*, edited by Philip P. Wiener, University of Virginia Library, http://xtf.lib.virginia.edu/xtf/view?docId=DicHist/uvaGenText/tei/DicHist2.xml;chunk.id=dv2-31;toc.depth=1;toc.id=dv2-31;brand=default.

Murrow, Gail, and Richard Murrow. "A Hypothetical Neurological Association between Dehumanization and Human Rights Abuses." *Journal of Law and the Biosciences*, vol. 2, no. 2, June 2015, pp. 336–64, https://doi.org/10.1093/jlb/lsv015.

Naddaff, Ramona, and Katharine Wallerstein. “The Wrong Words in the Wrong Times.” *Common Knowledge*, vol. 23, no. 1, 2017, pp. 91–100., doi:10.1215/0961754X-3692224.

NBCDFW. “Lt. Gov. Patrick Wants to Remove Tenure, Ban CRT in State Colleges, Universities.” *NBCDWF*. 18th Feb. 2022, https://www.nbcdfw.com/news/local/texas-news/coming-up-dan-patrick-proposes-plan-to-ban-critical-race-theory-in-state-schools/2894264/, Accessed 12th March 2022.

Nieuwburg, Elisabeth G. I., et al. “Emotion Recognition in Nonhuman Primates: How Experimental Research Can Contribute to a Better Understanding of Underlying Mechanisms.” *Neuroscience & Biobehavioral Review*, vol. 123, Apr. 2021, pp. 24–47.

O’Rourke, K. C. *John Stuart Mill and Freedom of Expression: The Genesis of a Theory*. Routledge, 2001.

Paresky, Pamela. “When Is Speech Violence and What’s the Real Harm?” *Psychology Today*, 4 August 2017, https://www.psychologytoday.com/us/blog/happiness-and-the-pursuit-leadership/201708/when-is-speech-violence-and-what-s-the-real-harm, Accessed 12 February 2022.

Parker, Richard. “Clear and Present Danger Test.” *The First Amendment Encyclopedia*, 2009, https://www.mtsu.edu/first-amendment/article/898/clear-and-present-danger-test.

Patridge-Hicks, Sophie and Gillian Russo. “Student Pledges Lawsuit Against University Disputing Disciplinary Actions.” *The Observer*. 17th Jul. 2020, https://fordhamobserver.com/48725/news/student-pledges-lawsuit-against-university-disputing-disciplinary-actions/, Accessed 12th March 2022.

Pieterse A.L, et al. “Perceived Racism and Mental Health among Black American Adults: A Meta-Analytic Review.” *Journal of Counseling Psychology*, vol. 59, no. 1, 2012, pp. 1–9., doi:10.1037/a0026208.

Premack, David. “The Infant’s Theory of Self-Propelled Objects.” *Cognition*, vol. 36, 1990, pp. 1–16.

Prieto, Julian. “Violence helped ensure safety of students.” *The Daily Californian.* 7 Feb. 2017, https://www.dailycal.org/2017/02/07/violence-helped-ensure-safety-students/, Accessed 9th of March 2022.

Richard, John R. “Freedom of Expression in the Digital Age: A Historian’s Perspective.” *Church, Communication and Culture*, vol. 4, no. 1, 2019, pp. 25–38.

Ricoeur, Paul. “Violence and Language.” *Journal of French and Francophone Philosophy*, vol. 10, no. 2, 1998, pp. 32–41., doi:10.5195/JFFP.1998.410.

Roberts, Joseph W. “Neutrality, Speech.” *The First Amendment Encyclopedia*, 2009, https://www.mtsu.edu/first-amendment/article/1003/neutrality-speech.

Roginsky, Alexandra B., and Alexander Tsesis. “Hate Speech, Volition, and Neurology.” *Journal of Law and the Biosciences*, vol. 3, no. 1, 2015, pp. 174–77.

Rohleder, Nicolas. “Stress and Inflammation - the Need to Address the Gap in the Transition between Acute and Chronic Stress Effects.” *Psychoneuroendocrinology*, vol. 105, 2019, pp. 164–171., doi:10.1016/j.psyneuen.2019.02.021.

Saxonhouse, Arlene. *Free Speech and Democracy in Ancient Athens*. Cambridge University Press, 2006.

Shick, Finnegan. “Erika Christakis leaves teaching role.” *Yale News*. 7th Dec. 2015, https://yaledailynews.com/blog/2015/12/07/erika-christakis-to-end-teaching/, Accessed 12th of March 2022.

Schmidt, Samantha. “Professor fired after defending blacks-only event to Fox News. ‘I was publicly lynched,’ she says.” *The Washington Post*. 26th Jun. 2017, https://www.washingtonpost.com/news/morning-mix/wp/2017/06/26/professor-fired-after-defending-blacks-only-event-on-fox-news-i-was-publicly-lynched-she-says/, Accessed 12th March 2022.

Shoaib, Alia. “A UCLA professor suspended in a row over grades for Black students claims it was to distract from the business school that’s ‘inhospitable’ to people of color.” *Insider*. 16th Oct. 2021, https://www.insider.com/ucla-prof-feud-with-anderson-school-management-over-black-grades-2021-10#:~:text=A%20UCLA%20professor%20was%20suspended,well%2Dtimed%20publicity%20stunt.%22, Accessed 12th March 2022.

Schuessler, Jennifer. “A Blackface ‘Othello’ Shocks, and a Professor Steps Back From Class.” *New York Times.* 15th Oct. 2021, https://www.nytimes.com/2021/10/15/arts/music/othello-blackface-bright-sheng.html, Accessed 12th March 2022.

Simon, Jeremy C., and Jennifer N. Gutsell. “Recognizing Humanity: Dehumanization Predicts Neural Mirroring and Empathic Accuracy in Face-to-Face Interactions.” *Social Cognitive and Affective Neuroscience*, vol. 16, 2021, pp. 463–73.

Skaaning, Svend-Erik, and Suthan Krishnarajan. "Who Cares About Free Speech? Findings from a Global Survey of Support for Free Speech." Justicia, May 2021, https://middelfartavisen.dk/wp-content/uploads/report_who-cares-about-free-speech_21052021.pdf, Accessed 6th of April 2022.

Skelding, Conor and Kerry J. Byrne. “Fordham U.prof fired after mixing up two black students in class.” *New York Post.* 11th Dec. 2021,

https://nypost.com/2021/12/11/fordham-prof-fired-for-confusing-two-black-students-in-class/, Accessed 12th March 2022.

Soave, Robby. "A Medical Student Questioned Microaggressions. UVA Branded Him a Threat and Banished Him from Campus." *Reason*. 4th Jul. 2021, https://reason.com/2021/04/07/microaggressions-uva-student-kieran-bhattacharya-threat/, Accessed 12th March 2022.

—. "Colorado State University Sign Directs Students 'Affected By a Free Speech Event' To Seek Help." *Reason.* 31st Feb. 2022, https://reason.com/2022/01/31/colorado-state-university-sign-affected-by-a-free-speech-event/, Accessed 12th of March 2022.

Soral, Wiktor, et al. "Exposure to Hate Speech Increases Prejudice through Desensitization." *Aggressive Behavior*, vol. 44, no. 2, 2017, pp. 136–46.

Stewart-Williams, Steve. *The Ape That Understood the Universe: How the Mind and Culture Evolve*. Cambridge University Press, 2018.

Stanger, Allison. "Understanding the Angry Mob at Middlebury That Gave Me a Concussion." *New York Times.* 13th Mar. 2017, https://www.nytimes.com/2017/03/13/opinion/understanding-the-angry-mob-that-gave-me-a-concussion.html, Accessed 12th March 2022.

Thompson, B, et al. "Complex Cognitive Algorithms Preserved by Selective Social Learning in Experimental Populations." *Science (New York, N.y.)*, vol. 376, no. 6588, 2022, pp. 95–98., doi:10.1126/science.abn0915.

Thucydides. *History of the Peloponnesian War*. Translated by Benjamin Jowett, Prometheus Books, 1998.

Thuman, Leo. “Stanford prof publicly shamed, censored for using racial slur in academic context.” *Campus Reform*. 11th May. 2020, https://www.campusreform.org/?ID=14853, Accessed 12th March 2022.

University of Chicago. “Report of the Committee on Freedom of Expression.” 2014, https://provost.uchicago.edu/sites/default/files/documents/reports/FOECommitteeReport.pdf, Accessed 12th March 2022.

Villasenor, John. “Views among college students regarding the First Amendment: Results from a new survey.” *Brookings*. 18th Sept. 2017, https://www.brookings.edu/blog/fixgov/2017/09/18/views-among-college-students-regarding-the-first-amendment-results-from-a-new-survey/, Accessed 9th of March 2022.

Waldron, Jeremy. *The Harm in Hate Speech*. Harvard University Press, 2012.

Wallace, Robert W. “The Power to Speak--and Not to Listen--in Ancient Athens.” *Free Speech in Classical Antiquity*, edited by Ineke Sluiter and Ralph M. Rosen, Penn-Leiden Colloquium on Ancient Values, 2004.

Ward, Jamie. *The Student’s Guide to Social Neuroscience*. 2nd ed., Routledge, 2017.

Xenophon. *Xenophon: Scripta Minora*. Translated by E.C. Marchant and G.W. Bowersock, Harvard University Press & William Heinemann, 1968.

Zimmer, Carl. “E. O. Wilson, a Pioneer of Evolutionary Biology, Dies at 92.” *The New York Times*. 27th Dec 2021, https://www.nytimes.com/2021/12/27/science/eo-wilson-dead.html, Accessed 19th March 2022.

Works Consulted

Alex, Daniel. "Speech Locked Up: John Locke, Liberalism and the Regulation of Speech." *Law School Student Scholarship*, vol. 154, 2013, https://scholarship.shu.edu/student_scholarship/154.

Brown, Paul. "Scientist 'killed Amazon Indians to test race theory.'" *The Guardian*. 23rd Sep. 2000, https://www.theguardian.com/world/2000/sep/23/paulbrown#:~:text=Thousands%20of%20South%20American%20indians,a%20book%20out%20next%20month., Accessed 19th March 2022.

Carey, Benedict. "Criticism of a Gender Theory, and a Scientist Under Siege." *The New York Times*. 21st Aug. 2007, https://www.nytimes.com/2007/08/21/health/psychology/21gender.html, Accessed 19th March 2022.

Gibson, Tobias T. "Bad Tendency Test." *The First Amendment Encyclopedia*, 2009, https://www.mtsu.edu/first-amendment/article/893/bad-tendency-test.

Jiang, Yanping, et al. "Basal Cortisol, Cortisol Reactivity, and Telomere Length: A Systematic Review and Meta-Analysis." *Psychoneuroendocrinology*, vol. 103, 2019, pp. 163–172.

John, Richard R. "Freedom of Expression in the Digital Age: A Historian's Perspective." *Church, Communication and Culture*, vol. 4, no. 1, 2019, pp. 25–38.

Murphy, Andrew R. *Conscience and Community: Revisiting Toleration and Religious Dissent in Early Modern England and America*. Pennsylvania State University Press, 2001.

O'Rourke, K. C. *John Stuart Mill and Freedom of Expression: The Genesis of a Theory*. Routledge, 2001.

Parker, Richard. “Clear and Present Danger Test.” The First Amendment Encyclopedia, 2009, https://www.mtsu.edu/first-amendment/article/898/clear-and-present-danger-test.

Rohleder, Nicolas. “Stress and Inflammation – the Need to Address the Gap in the Transition between Acute and Chronic Stress Effects.” *Psychoneuroendocrinology*, vol. 105, 2019.

Strossen, Nadine. *HATE: Why We Should Resist It with Free Speech, Not Censorship*. Oxford University Press, 2018.

APPENDIX: DePauw University

In both 2020 and 2021 DePauw University was crowned as the worst university for free speech out of first out of 55 and then out of 154 US institutions according to the survey sponsored by College Pulse, the Foundation for Individual Rights in Education (FIRE), and RealClearEducation. The survey covers three kinds of questions across which DePauw ranked poorly. The first kind of questions rates students' comfort with exercising their own right to free expression by expressing their views on a controversial political topic in various settings. DePauw's ratings reveal a concerning atmosphere of self-censorship with 31% of students feeling very uncomfortable expressing their views on a controversial political topic in a classroom and 20% very uncomfortable expressing their views to other students in common areas (College Pulse/FIRE). The second kind of questions explores the students' tendencies to permit others to exercise their right to free expression. Questions were asked about whether you would be supportive of allowing speakers with various controversial positions of speaking and how permissive you are towards actions blocking people from exercising their free expression. DePauw's students foster such censorious tendencies as 30% see it as at least sometimes acceptable to use violence to stop a campus speech (College Pulse/FIRE). The third kind of questions aimed to assess the students' view of how much their administration protects freedom of expression. DePauw students rated their administration exceedingly poorly as only 2% of students found it extremely clear that their administration protects free expression and only 1% of students thought it extremely likely that the administration would defend a speaker's right to express their views should a controversy occur (College Pulse/FIRE).

An environment of self-censorship and over-sensitivity is very difficult to repair, but an important aspect of making DePauw's campus a better place for a free exchange of ideas is

for the administration to clearly communicate that they defend the value of free expression. In an effort to do so, DePauw is hoping to adopt its own statement on the university's values regarding freedom of expression by the fast approaching May 2022. DePauw's decision not to adopt the statement of the University of Chicago who ranked highest in the surveys in 2020 and second highest in 2021 is in itself a peculiar one. Joining the ranks of schools who have adopted the Chicago Principles would be a great way for DePauw to make a public commitment to improving its campus climate and addressing its free speech problem so that DePauw's students can enjoy a comparable learning environment to that of the nation's finest institutions. In fact, the ways in which a draft of DePauw's statement diverges from the Chicago Principles reveals fascinating clues into how playing into the "speech equals violence" fallacy with loose diction obstructs well-meaning aspirations to defend freedom of expression.

The draft of DePauw's statement that has not yet been approved and is subject to potential changes includes several important similarities with the Chicago statements. It aims to emulate Chicago Principles in expressing faith in individual autonomy by asserting that it "is not the role of the University to shield individuals from ideas or opinions they find unwelcome, disagreeable, or even offensive" and that "it is for the individual members of the community, not for the institution, to make judgments for themselves and act on those judgments, not by seeking to suppress speech but by openly contesting ideas they oppose" (DePauw). However, the draft also asserts that one of the key requirements of DePauw's Mission is to foster an "environment in which everyone feels safe to present their diverse ideas, even those that may seem offensive or repulsive" (DePauw). While it is certainly important for students to *be* safe and protected when voicing diverse ideas, requiring that they *feel* safe at all times conflicts with the significance of free expression for education—in

discussion and debate students ought to be brave and exploratory even when it feels risky or as University of Chicago's president asserts: "education should not be intended to make people comfortable, it is meant to make them think" (University of Chicago). Similar to conflating being and feeling safe, the draft of DePauw's statement imprecisely endorses the idea that speech equals violence by stating: "we must be ever mindful of the powerful duality of words. They can wound as well as illuminate and teach" (DePauw). To reiterate my arguments from earlier chapters, it is an imprecise and a dangerous conflation of figurative and literal expression to say that words wound. Indeed, expression can *lead to* emotional distress and psychological discomfort. *Using* certain words can lead to distress and discomfort. But distress and discomfort differ in kind from a wound or an injury.

The same kind of a seemingly innocent lapse of linguistic precision occurs in the draft's section discussing "Community Expectations and Responsibilities" when the authors express the hope that members of DePauw community "will consider carefully the use of words that may *harm* [emphasis added] others" (DePauw). Such instances of poor wording create a sense of conflict within the very draft of DePauw's statement. Namely, it appears as if DePauw endorses the erroneous notions that speech equals violence, words wound, and expression harms and yet the draft states that the university aspires to an environment in which disagreement is not perceived as a threat: "In a community marked by true inclusion and equity, *even fierce debates* [emphasis added] about a range of differences of opinions and perspectives *are not experienced as personal attacks* [emphasis added] on one's very humanity and sense of well-being and belonging" (DePauw). On the one hand, it seems as if this sentence forwards the very important notion that expression in an educational setting should not be experienced as a threat to one's humanity, well-being, and belonging. On the other hand, however, the sentence could mean to assert that students' subjective experience of

speech is a marker of "true inclusion and equity." The latter would mean that the draft's authors are saying that when DePauw's students take expression as a personal attack on their humanity, they are not over-reacting and perceiving speech as more powerful than it is. Rather, their upset acts as an indicator that DePauw's administration ought to work harder in ensuring "true inclusion and equity" is attained. The future of free expression at DePauw is bleak if its administration presents "inclusion and equity" as a precondition for free expression rather than the other way around. A healthy culture of free expression is necessary for members of DePauw's community to make sense of the meaning of the vague values of inclusion and equity and realize them in practice: free expression is a precondition for enacting those values rather than a reward after they are attained.

Compared to the Chicago Principles, the draft of DePauw's statement seems to present freedom of expression as much more dependent on competing values. The Chicago Statement emphasizes that while civility and mutual respect are important "concerns about civility and mutual respect can never be used as a justification for closing off discussion of ideas" (University of Chicago). DePauw's draft, on the other hand warns that "legal protections for free expression /…/ may sometimes supersede the values of civility and mutual respect" and instead of focusing on defending freedom of expression even in controversial cases suggests it should be exercised cautiously: "we encourage members of the community to consider these values carefully in exercising their fundamental right to free expression" (DePauw). Pressing community members to carefully consider competing values in exercising their freedom of expression seems admirable, but it is counterproductive to addressing the culture of self-censorship: too great a concern over how expression will be perceived is likely to play a role in currently stifling DePauw students' willingness to express their views inside and outside of the classroom. Respect and civility are incredibly important,

but what if a culture of politicized oversensitivity takes some kinds of views to be harmful, dehumanizing, and wounding no matter how carefully and respectfully they are expressed?

DePauw University is not unique in struggling to find room for free expression while catering to the trends of political correctness and extreme sensitivity to social justice issues. Catering to the customer by adopting the rhetoric supportive of the ideas that language, expression, and speech can equal violence, however, leaves little to no room for freedom of expression to exist in a meaningful way. While the efforts from DePauw's administration to address the worrying state of freedom of expression on its campus are an encouraging step forward, they ought to begin by voicing a firmer and clearer endorsement of the value than the current draft provides. Most importantly, the administration ought to commit to defending their faculty, staff, and students when exercising their right sparks discontent. Fostering an environment of civility, respect, inclusion, and equity is important, but those values must not be used as an excuse to perpetuate a soft stigma currently placed on freely, openly, and daringly expressing ideas at DePauw University.