

# **From “More Speech” To Enhanced Counterspeech: Towards a State Oriented Solution**

By  
Can Kayaş

Submitted to the Department of Philosophy, Central European University

In partial fulfillment of the requirements for the degree of Masters of  
Arts

Supervisor: Andres Moles

Budapest, Hungary

2020

## Contents

<b>Introduction .....</b>	<b>3</b>
<b>Categories of Hate Speech .....</b>	<b>3</b>
<b>Counterspeech and its Relation to Free Speech Principles.....</b>	<b>6</b>
<b>Marketplace, Truth and Unsubstantiated Counterspeech .....</b>	<b>12</b>
<b>Unsubstantiated Counterspeech Claim in Relation to the Marketplace of Ideas.....</b>	<b>13</b>
<b>Empirical Challenges to the Truth Claim of the Marketplace Metaphor.....</b>	<b>14</b>
<b>Counterspeech In Practice.....</b>	<b>20</b>
<b>Practical Concerns About Unsubstantiated Counterspeech as a Speech Act .....</b>	<b>22</b>
<b>Epistemic Injustice in the Context of Counterspeech .....</b>	<b>25</b>
<b>A Level Playing Field .....</b>	<b>29</b>
<b>Personal Reasons Against Engaging in Unsubstantiated Counterspeech .....</b>	<b>30</b>
<b>Counterspeech Enhanced .....</b>	<b>33</b>
<b>Initial Limitations on State Counterspeech.....</b>	<b>33</b>
<b>Two Further Limits on State Counterspeech.....</b>	<b>37</b>
<b>Conclusion.....</b>	<b>53</b>

## From “More Speech” To Enhanced Counterspeech: Towards a State Oriented Solution

### Abstract

The phrase that the best answer to bad speech is good speech has been around for quite a while in popular rhetoric against hate speech restrictions. On the rare occasions where speech is permissibly restricted, such as the instances of defamation, the questions about how the offense should be redressed (and to a lesser extent who should redress it) are important and demand substantive answers. I argue that such due diligence is absent in the traditional philosophical and legal approaches to answering back where the speech *is* protected, but the speakers are inflicting serious harms on their targets. Not only the relevant questions remain unasked but also the answers are assumed. In the following pages, I will be highlighting philosophical and sociological views that are attentive to the demands and specifics of counterspeech. I will also be unpacking the implications behind and testing the viability of solutions that are assumed to be effective by legal authorities and philosophical commentators. My conclusion is that the empirical grounds for the claim that “more speech” is the best remedy is highly contested while its moral implications are often unacceptable. A new approach to counterspeech must be found if we aim to prevent and redress the harms as well as transform the speakers without resorting to the blunt instrument of censorship. I will be offering Brettschneider’s expressive state theory as a candidate approach and argue that it is compatible –and if not, can be made compatible– with a range of positions on free speech.

**Keywords:** free speech remedies; more speech; counterspeech; marketplace of ideas; expressive state

## **Introduction**

Throughout this work, I will exclusively focus on expressions that are generally recognized as hate speech. I argue for the moral urgency of a robust counterspeech based response –along with or in the absence of the protections of hate speech– against the harms of hate speech. I take counterspeech as any type of response that aims to heal the harms and undermine the adverse effects of hate speech. Most of what has been said about the benefits of free debate apply to counterspeech since both terms cover roughly the same area. With that in mind, I also intend to demonstrate that counterspeech may have a range of application that extends beyond the traditional conception of a “town hall debate.” Counterspeech as an act could be verbal or symbolic; it can be immediately personal as in a face-to-face response; personal but mediated as in a personal letter to the offender; or it can be impersonal like an opinion piece in a newspaper about a general subject such as combatting racism. The moral demandingness, effectiveness, and the locus of responsibility may vary significantly from case to case. And any serious counterspeech theory should ask these preliminary questions: Who should answer back? Are they capable? Is their effort worthwhile in terms of the results that they expect to achieve? These are the initial necessary steps of specification and elucidation that are sadly absent from the well-known, traditional ideas about answering back to hate speech.

## **Categories of Hate Speech**

The notion of hate speech that I will be operating with, is minimally controversial, though as every definition of hate speech runs the risk of being incomplete. I will follow a broad definition of hate speech which includes not only the verbal communications but also the symbols and images that aim to cause injury to or discriminate against others based on their membership to a certain group. Examples to these groups include but are not limited to one’s nationality, religion, race, and sexual orientation. Note that, what makes a cutting or

discriminatory expression hateful, is not the emotional state of its speaker but its content, and the fact that it does so on a morally arbitrary footing. By focusing on hate speech, I intend to leave out some forms of expressions that fall into the category of harmful speech such as factual disinformation, defamatory speech, and false advertising. These issues may also need a counterspeech remedy, but I think they deserve their own, distinct discussions.

It will be worth unpacking what types of speech acts I mean by the general term “hate speech,” since one’s stance regarding one category may not reflect his stance on the other. One categorization that readily comes to mind treats all speech acts as either high or low-value where high-value speech has a stronger claim for protection. Broadly speaking, low-value speech (e.g. incoherent shouting, provocative insults as well as defamation and false advertising) is the speech that contributes little to nothing to free speech interests (i.e. the reasons for which we defend protections for speech). This is a traditional legal method that makes regulatory decisions concerning speech a bit more sensitive to the type of the message and it can be used specifically for instances of hate speech. Whether a form of speech is considered high or low value will certainly impact considerations regarding counterspeech, but I think there is still need for further specification. In this regard, I find Caleb Yong’s categorization helpful which distinguishes forms of hate speech as targeted and diffuse vilification, political hate speech and hateful statements of fact. Here the contours of each category depend on the speakers’ intentions as well as the specific method that he or she chooses to convey her message (if there is one). In this scheme insults and what U.S. courts generally classify as fighting words would fall under targeted vilification where the speaker intends to hurt or intimidate a person or a small group because of their race, gender, etc. They often occur face-to-face but exceptions can be found such as threatening letters, online bullying, and so on. I take that such expressions to convey negligible propositional content and only tangentially relate to reasons for protecting speech. Therefore, in most cases, they

can be considered low-value speech, legally speaking, equal to loud ramblings of a madman. Either way, if the speaker intended to convey their political stance to convince others, he or she would be better suited to express their opinion through means that don't alienate the audience from the get-go. Intent to hurt and/or intimidate can also be expressed in symbolic forms such as wearing armbands, raising flags, burning crosses, or using a specific set of symbols or emojis in their online profiles in which case they can be classified as diffuse vilification. Even though its cognitive content is often obscure and limited, when it is political, diffuse vilification provokes a more urgent counterspeech response compared to targeted vilification which usually triggers anger or apathy. Political expressions that can be considered hate speech call for discrimination against, disenfranchisement of or even extermination of a certain group of people, are sometimes considered protected speech and are in urgent need of a counterspeech response (as I will point out below they need equally strong counterspeech response even if they are unprotected). The intent behind those expressions does not solely lie in hurting or intimidating their targets and their content is not simplistic. They are often considered high value, albeit extremely harmful speech. Lastly, the expressions which take the form of personal statements of value or purported fact can be considered hate speech when their goals align with the goals that I've outlined above regarding political hate speech. Contrarily here, the speaker's intent may not involve convincing others to a certain political agenda, it may be as simple as expressing a deeply valued belief, even though it will cause distress for others. When we are considering speaking in response to any single one of these categories, our methods, our ability to respond, and the goals we aim to reach through our response will vary from case to case. One's main concern when answering to a racial epithet may be their own safety, they may choose to keep a cool head and walk away or they may give in to their anger, or they may choose to engage in a fruitless effort to somehow convince the offender; whereas dealing with a dispassionate and

methodical discriminatory political campaign calls for an inherently different counterspeech approach. Any response-based solution to the harms of speech must take heed of the nuances between and exigencies of different types of hate speech.

### Counterspeech and its Relation to Free Speech Principles

Another important subject that I need to clarify before moving on to counterspeech relates to the limits of free speech. Traditionally, counterspeech has been presented in a dichotomic relation to criminal sanctions against bad speech as the only permissible remedy where bans are impermissible. We can see remnants of such views as far back as Mill:

That 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 ... He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right. *These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise.*<sup>1</sup>

But the choice is not between restricting speech or permitting bad speech and speaking back.

When one side advances, the other does not have to retreat. Recent work on the subject of counterspeech provides examples of philosophers who draw narrower limits on protected speech but still support a strong counterspeech approach when the bans are ineffective; or philosophers who believe that some speech does not deserve protection but still argue in favor of a strong, complementary notion of counterspeech.<sup>2</sup> Additionally, even though restrictions

---

<sup>1</sup> John Stuart Mill, *Essays on Politics and Society*, Collected Works of John Stuart Mill, v. 18-19 (Toronto ; Buffalo: University of Toronto Press, 1977), 224. My italics.

<sup>2</sup> Caleb Yong, “Does Freedom of Speech Include Hate Speech?,” *Res Publica* 17, no. 4 (July 13, 2011): 385, <https://doi.org/10.1007/s11158-011-9158-y>; Katharine Gelber, “Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia),” in *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, ed. Michael Herz and Peter Molnar (Cambridge: Cambridge University Press, 2012), 198–216, <https://doi.org/10.1017/CBO9781139042871.016>; Jeffrey W. Howard, “Terror, Hate and the Demands of Counter-Speech,” *British Journal of Political Science*, 2019, 1–16, <https://doi.org/10.1017/S000712341900053X>; Mary Kate McGowan, “Responding to Harmful Speech : The More Speech Response, Counter Speech, and the Complexity of Language Use,” in *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public*, ed. Casey Rebecca Johnson, 2018, <https://doi.org/10.4324/9781315181189-12>; R. Langton, “Blocking as Counter-Speech,” in *New Work on Speech Acts*, ed. Daniel Harris, Daniel Fogal, and Matt Moss (New York: Oxford University Press, 2017), 144–64, <https://doi.org/10.1093/oso/9780198738831.003.0006>.

and counterspeech are generally presented in an interdependent fashion, the point of counterspeech and the point of free speech justifications is quite distinct. Counterspeech aims to prevent or redress the dangers and harms of bad speech with the additional goal of changing the speaker's mind –I will be referring to those as reparative and transformative goals of counterspeech respectively. On the other hand, free speech justification strategies aim to defend values such as autonomy, democratic participation, and truth and try to define the limits of protected speech in a compatible way with those values. Counterspeech claims do not require taking a stance within the debate about speech regulation. Indeed, because of the compatibility of counterspeech with different positions on the regulation of speech I will take care to present my view in a form that can be potentially useful to and compatible with all of them. Still, at this point, the relation between the aims of counterspeech and free speech interests is mysterious, and I would like to pause on this subject before I begin with the thesis proper.

Throughout the rest of the thesis, I'll be operating with two main goals for counterspeech in mind. Progress towards these goals is the subject of social, empirical research and their pursuit can be described as a consequentialist project, meaning that counterspeech succeeds insofar as it furthers these aims. First is the transformative goal in which counter-speakers attempt to change the minds of hate speakers to whom they are responding to or the minds of the indirect audience. This process can be post factum where the transformative act is aimed towards people who already hold and express hateful views or it can be ante factum/pre-emptive in the form of educating or advising the general populace who will be exposed to hateful speech or arguments in later stages of their life. Ideally, success towards the transformative goal will reduce the amount and salience of hateful expressions within the public debate and ordinary life thereby blocking the harm of future, potential instances of hate speech. The second goal of counterspeech would be reparative



where the counter-speaker attempts to counterbalance the harm that is done by the hateful expressions. There is an important difference between counterbalancing and negating harm, but the complete negation harm would be an insurmountable challenge for counterspeech since negation entails identity or near identity between post-harm and pre-harm state of the victims. On the other hand, counterbalancing requires proportional compensation for the harm. As a concrete example, increased government funding for statues of minority historical figures would be an attempt at counterbalancing the harms of hate speech and can be considered as progress towards the reparative goal of counterspeech. The government can't undo the offense and distress caused by commissioning or maintaining statues dedicated to racist historical figures but can try to make amends by promoting the idea of equal public visibility and representation of hitherto unjustly ignored or historically sidelined people.

Though the overarching aim of counterspeech is the reduction of harm –be it through reparation, transformation, or a combination thereof– it can also have secondary effects that are unrelated to harm. Namely, it can further some free speech interests that we appeal to when justifying protections for speech such as democratic participation and autonomy/self-nourishment of individuals. I take these secondary effects as neither necessary nor sufficient for the success of counterspeech, meaning that a theory of counterspeech, which promises contributions to a flourishing democratic process or allows people to make their minds more safely and efficiently without significant reduction in harms of speech, would not pass the test.

Prudence demands that I specify exactly what I mean by harms of hate speech, especially since I'm taking the reduction of these harms as the overarching aim of the counterspeech. The most direct and clinically well-proven harm of hate speech is the significant physical and mental distress it causes for its victims. Being the target of discriminatory speech acts is associated with increased depression, drug consumption, and

suicide rates as well as declining physical health among ethnic minorities and the LGBT community.<sup>3</sup> As we move towards societal harm caused by hate speech, the directness of the relation between cause and effect becomes less clear. However, societal harm is no less significant than the ones I previously mentioned. These harms mainly consist of people, small groups, or even masses being persuaded by racist, xenophobic, misogynistic, or patriarchal beliefs as well as falling victim to incitement to discriminatory violence and this is the area where we need an effective counterspeech solution the most. One philosopher who is skeptical about such an approach in favor of more regulations about hate speech is Jeremy Waldron. And although his stance toward counterspeech does not align with mine, his further specification on a certain type of societal harm is worth mentioning. Waldron, among others, emphasizes what he calls dignitary harms where the victims of discrimination are deprived of their right to be perceived as equal citizens by hate speakers and governments that protect hate speech.<sup>4</sup> Debates around the regulation of hate speech as well as the ones about the exact nature and origin of the harm of speech also often recognize the significance and salience of these harms. However, I will not be engaging with the subject of causation and constitution of harm throughout the thesis. I'm sure, by their own, physical, and mental harms that are proven to be caused by hate speech warrant more attention allocated to counterspeech. But whether the sum of the social, mental, and physical harms justifies broad regulations against hate speech is a different matter which deserves at least some preliminary remarks.

Significance of the above harms still in mind, harm reduction is neither the sole nor the primary concern of the society with regards to free speech. There are good reasons against

---

<sup>3</sup> Gilbert C. Gee, "A Multilevel Analysis of the Relationship Between Institutional and Individual Racial Discrimination and Health Status," *American Journal of Public Health* 92, no. 4 (April 2002): 615–23; Ian H. Meyer, "Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence," *Psychological Bulletin* 129, no. 5 (September 2003): 674–97, <https://doi.org/10.1037/0033-2909.129.5.674>; Mari J. Matsuda, *Words That Wound: Critical Race Theory, Assaultive Speech, And The First Amendment*, 1 edition (Boulder, Colo: Westview Press, 1993).

<sup>4</sup> Jeremy Waldron, *The Harm in Hate Speech* (Cambridge, Massachusetts London, England: Harvard University Press, 2014), 5.

treating speech cases according to a general harm principle, where government interference is limited only barring the cases in which it can prevent harm to others. The free speech interests that are at stake are often significant enough to allow protections for hate speech *despite* its harms. Indeed, there is great value in everyone having a fair chance of participating in the democratic governance process, despite the harms of some of those views. And there is great value in people having access to the widest possible set of information while making decisions that will shape the trajectory of their lives, even though it means they will be exposed to outrageous and vile opinions of others in the process. However important they may be, these speech interests are not absolute and the matter of how much harm that we can tolerate or what kind of dangerous ideas we can allow as a society before these interests themselves become meaningless is an open question. The absolute and urgent fact is that these harms and dangers demand an effective counterspeech response on the part of governments, civil society, and concerned citizens.

In the first section, I engage a particular, *unsubstantiated* understanding of counterspeech which simply suggests that the *best remedy to bad speech is more speech*. I point to the results of natural and social sciences that put the question on the validity of this claim as well as the associated metaphor of the marketplace of ideas. In the second section, I continue with the epistemic and linguistic concerns about unsubstantiated counterspeech. These are useful in revealing that unsubstantiated counterspeech unfairly burdens the targets of hate speech in the absence of additional governmental remedies. Lastly, I focus on the possibility of a governmental counterspeech proposed by Corey Brettschneider,<sup>5</sup> which is a significant improvement over unsubstantiated counterspeech. The expressive state theory may run afoul of some liberal commitments against unjustified coercion and for vibrant civil

---

<sup>5</sup> Corey Brettschneider, “When the State Speaks, What Should It Say? The Dilemmas of Freedom of Expression and Democratic Persuasion,” *Perspectives on Politics* 8, no. 4 (December 2010): 1005–19, <https://doi.org/10.1017/S1537592710003154>.

society. I leave open the possibility of adoption of state counterspeech by non-liberal scholars or the liberals who have weaker commitments to certain principles. I also argue that nothing is holding traditional liberals back from modifying and reducing the scope of Brettschneider's framework and using it (even in a narrower form) as a good substitute to unsubstantiated counterspeech.

## Marketplace, Truth and Unsubstantiated Counterspeech

With the limits of my discussion thus sketched I would like to move on to the subject of unsubstantiated counterspeech, which has found itself a central role both in legal decisions and political rhetoric but remained theoretically unsubstantiated until recently. It is unsubstantiated because when we encounter it in traditional legal texts, the effectiveness of this remedy and its possible moral ramifications are presented with little or no argument. Such an armchair conception of counterspeech has been famously voiced by Justice Brandeis of the U.S Supreme court. In his concurring opinion to *Whitney v. California* he states that “the remedy to be applied” to the unwanted effects of speech “is more speech, not enforced silence.”<sup>6</sup> Along similar lines, Justice Holmes stated in his dissenting opinion in *Abrams v. the United States* that “the ultimate good desired is better reached by free trade in ideas” and “that the best test of truth is the power of the thought to get itself accepted in the competition of the market.”<sup>7</sup> With regards to the specific context of hate speech, ACLU states that “where racist, misogynist, homophobic, and transphobic speech is concerned, the ACLU believes that more speech –not less– is the answer most consistent with our constitutional values.”<sup>8</sup> The statements above can be historically traced back to Milton’s *Areopagitica* and J.S. Mill’s detailed arguments in favor of diverse, unrestricted debate, advanced in the second chapter of *On Liberty*. In popular political discourse, we encounter similar claims in the form of arguments in favor of a *free marketplace of ideas*. However, the marketplace metaphor comes with significant baggage of empirical suppositions which lead to troubling implications. I will begin by unpacking and testing some of the empirical suppositions and conclude that the

---

<sup>6</sup> *Whitney v. California*, 274 U.S. 357 (1927).

<sup>7</sup> *Abrams v. United States*, 250 U.S. 616 (1919).

<sup>8</sup> “Speech on Campus,” American Civil Liberties Union, accessed June 8, 2020, <https://www.aclu.org/other/speech-campus>.

empirical challenges against what I call *unsubstantiated counterspeech* significantly hurt its chances to be the best remedy to bad speech.

### **Unsubstantiated Counterspeech Claim in Relation to the Marketplace of Ideas**

The claim in favor of unsubstantiated counterspeech is a familiar one, reflected in part in the influential public statements and court decisions I've cited above. It stipulates that there are harms that come as a consequence of bad speech and *more speech is better in ameliorating those harms compared to government restrictions*. On the other hand, the marketplace metaphor suggests that in a free and unrestricted public debate, truthful statements have a strong tendency to eventually win popular support over falsehoods. I take this unsubstantiated counterspeech claim to be an integral part of the famous “marketplace of ideas” metaphor and I believe the close relationship between the two will allow me to use empirical arguments against the latter as criticism against the prior.

Firstly, let me link the unsubstantiated counterspeech and the marketplace metaphor more explicitly: Success of counterspeech is necessary for the marketplace of ideas to yield desired results –results that justify its viability as a solution to bad speech. But I must also note that the ends of counterspeech and the marketplace metaphor do not completely overlap. First aims for the reduction of harm while the latter –in its traditional formulation– aims to reduce the number of falsehoods. Still, some versions of the marketplace metaphor indeed allow for a harmonious reading. For instance, the claim that free debate will convince others and heal the harms of bad speech seems compatible with the gist of the marketplace metaphor. Operating with such a reading can allow the goals of counterspeech and the marketplace to converge on the same values. Alternatively, we can work with examples of hate speech that are truth-apt (examples of misinformation with discriminatory content such as “refugees commit more crimes” are abound). If such examples are objectively false and if

the marketplace is supposed to reveal and amend these falsehoods, then we can say that it furthers the aims of counterspeech to some extent. Of course, both of these approaches hinge on the empirical demonstrability of the marketplace's efficacy –a hardly uncontroversial matter as I will demonstrate below.

We can branch the claim about unsubstantiated counterspeech into strong and weak versions where the implausibly strong claim says that the counterspeech is the best remedy to bad speech –period; and the weak version only states that: *in comparison to censorship*, unsubstantiated counterspeech is better at redressing the harms of speech. The comparative aspect is explicit in Brandeis' opinion quoted above and it is the crucial point of Mill's argument against restricting speech.

### **Empirical Challenges to the Truth Claim of the Marketplace Metaphor**

Both the weaker and the stronger claims face significant empirical challenges both in general contexts and specifically in cases of hate speech. However, it is important to note that, in both versions, the content of the claim is more nuanced than the popular phrase: “the truth will always prevail in the marketplace of ideas.” Indeed, Mill explicitly warns that:

... the dictum that truth always triumphs over persecution is one of those pleasant falsehoods which men repeat after one another till they pass into commonplaces, but which all experience refutes ... It is a piece of idle sentimentality that truth, merely as truth, has any inherent power denied to error of prevailing against the dungeon and the stake.<sup>9</sup>

Mill's hope instead lies in the *persistence* of truth, that “it may be extinguished once, twice, or many times, but in the course of ages there will generally be found persons to rediscover it.”<sup>10</sup>

Presumably, as the people persist in answering back, truthful information will secure an undeniable and unquenchable place for itself. An additional implication is that when truthful and justice-promoting statements become popular ideas or even social norms, this should

<sup>9</sup> Mill, *Essays on Politics and Society*, 238.

<sup>10</sup> Ibid., at 239.

amount to a significant contribution to canceling and redressing the effects of harmful and false speech and even changing the minds of speakers of such speech. Mill's foresight certainly clears the air for future debates on this subject, but it also creates additional problems that I will point out shortly. At the end of this section, I intend to demonstrate that the empirical research shows that unsubstantiated counterspeech is ill-equipped to accomplish these goals and an enhanced reformulation is needed.

With this cautionary note, I first move on to the empirical challenges to unsubstantiated counterspeech, which consists of merely answering back to hate speech. The main problems with the empirical side of this argument are the lack of definitive data in support of it as well as the difficulty of obtaining such data. Ho & Schauer point out that vagueness about which particular truth-relation is in effect in the marketplace of ideas, generates empirical hurdles from the get-go since the range of plausible interpretations about truth is likely to determine the kind of test(s) that the claim will be subject to.<sup>11</sup> When the claim is taken to say that *truth of a statement has a positive causal effect on the likelihood of acceptance of statements* (this is a relatively restrained version of "pleasant falsehood" that the Mill was cautioning against), studies in social psychology provide plenty of evidence to the contrary. Though the claim above is the cashed-out version a particular understanding of marketplace metaphor, it can be reformulated as a counterspeech claim; i.e. counterspeech is a good remedy (or a better remedy than restrictions) since expressing the truth will increase the chances of it becoming widely accepted. Research shows that the identity of the speakers, their social status, their authority among their audience, and the audience's prior beliefs on divisive topics have a substantive impact on whether the audience accepts the information

---

<sup>11</sup> Daniel E Ho and Frederick Schauer, "Testing the Marketplace of Ideas," *New York University Law Review* 90 (2015): 1161–75.



presented.<sup>12</sup> This becomes especially troubling with regards to the effectiveness of counterspeech when the audience is presented with reliable, balanced information but chooses to reject or ignore it due to the reasons mentioned above. To be clear, none of the factors above by themselves or in combination disprove the causal effect of truth on the audience. What it rather shows is that certain epistemic phenomena, which factor-in regularly in our daily lives, have a significant chance to diminish or cancel this effect. Presuming that originators of the market metaphor had in mind a complex and dynamic system, real effects of which take place over generations and after multiple layers of deliberation, instead of a singular instance of examination by the public (like a referendum), it is very hard to isolate and test the actual causal effect of truth separately from the causal effect of contrary epistemic phenomena. Finally, the same claim, when formulated as the weak claim about counterspeech—that the truth has a stronger causal effect on the likelihood of acceptance of statements, therefore speaking back is a better remedy *in comparison to government restrictions*—is also subject to criticisms that I’ve outlined above. This is true even if, compared to unsubstantiated counterspeech, the alternatives for belief formation (such as government pre-selecting the available information and depriving the public of dissenting voices) are just as hard to empirically test and just as entangled with the psychological factors above. When the methods and criteria of the examination are unclear or out of reach, it is not possible to definitively claim which particular approach is *better*.<sup>13</sup>

---

<sup>12</sup> Richard E. Petty and John T. Cacioppo, *Attitudes And Persuasion: Classic And Contemporary Approaches* (Boulder, Colo: Routledge, 1996).

<sup>13</sup> Skepticism about empirical foundation of the marketplace metaphor may permeate candidate arguments in favor of counterspeech and may cast similar doubts on their validity. In other words, the research into mechanics of free debate is either insufficient or too complex to validate the claims of the marketplace metaphor, the same complexity detracts from any consequentialist counterspeech claim. However, the reasons for the skepticism about the success of the marketplace metaphor are its obscurity (mystery about the actual truth-effect the market aims to bring about) and its simplicity (wishful thinking that assumes away the empirical challenges). Therefore, I think that a fleshed out counterspeech claim with the help of empirical research moving into the right direction (as it slowly does especially in areas of combatting hate speech, fake news and conspiracy theories), can face the skeptical challenge above.

Alternatively, the general marketplace of ideas mechanic can be understood as *truth-eliciting*. Presumption voiced and challenged in Ho & Schauer is that the marketplace of ideas encourages people to come forward and share the truth. The claim goes that, *the market encourages people to supply the public with truthful propositions in reply to falsehoods, and that supply will be greater than it would have been under censorial practices*. Though this is a reasonable claim, there is research suggesting that emotional factors do compete with and can overwhelm the truth as a motivator for speaking back. Heath et. al. show that people are more likely to share urban legends when those legends are exciting a strong emotional response.<sup>14</sup> In their study, the stories that were more likely to be believed to be true were rated higher in terms of whether the participants would pass them along, compared to stories which mainly contained practical information or a moral lesson. However, stories that contained disgusting elements had also a high likelihood of being shared. When these factors were comparatively tested, and the levels of disgust and plausibility were manipulated, the study found that people were *most* likely to share very disgusting but highly implausible stories.<sup>15</sup> When their hypothesis was tested in an uncontrolled environment, by studying websites which cataloged urban legends, the results confirmed that “more disgusting legends are more successful in the social environment”<sup>16</sup>; those stories were shared more and, as a consequence, had a more persistent impact on public consciousness.

Overall, the study suggests that competition in the marketplace of ideas can be influenced by emotional factors, in some cases so much so that it detracts from or overrules the motivating power of plausibility. We can also take these results as a challenge to Mill’s hope for the persistence of truth, as the study shows that falsehoods may also have a

---

<sup>14</sup> Chip Heath, Chris Bell, and Emily Sternberg, “Emotional Selection in Memes: The Case of Urban Legends,” *Journal of Personality and Social Psychology* 81, no. 6 (2001): 1028–41, <https://doi.org/10.1037/0022-3514.81.6.1028>.

<sup>15</sup> *Ibid.*, at 1037.

<sup>16</sup> *Ibid.*, at 1039 .

successful and relatively long lifespan. Although these findings don't say much about the truth eliciting powers of a free market of ideas compared to government restrictions, the claim is vulnerable to the results above insofar as it is ignorant to the other strong motivations for (and the disincentives against) speaking back.

On this subject, Laura Beth Nielsen, in a study compiling and documenting the experiences of targets of sexually and racially discriminatory speech, states that “responses to both race-related and gender-related street speech are the product of complicated calculations made by the targets of such speech.”<sup>17</sup> According to the results, people generally remain reluctant to respond to hate speech, even after lengthy deliberations, due to various factors – but mainly for their concern for personal safety. Interviews with victims reveal that targets of hate speech usually are motivated to speak back and set things right, just as the claim above predicts. However, in an overwhelming majority of the cases, this motivation gets overridden because of the participants’ serious and genuine concern for their physical safety. Indeed, the fact that targets of hate speech lack the authority to speak back or may feel concerned about their safety –when faced with a hate speaker whose intent could not be further away from engaging in civil and factual debate– is an issue that not only this claim but also the main argument for “simply speaking back” overlooks. I will be returning to this subject later on in greater depth when discussing a more robust alternative to unsubstantiated counterspeech. For now, it is sufficient to say that the empirical support for the market’s ability to encourage people to speak up is inconclusive.

People may have various incentives when they offer their ideas to compete in the market and respond to bad speech, but the primacy of an overriding motivation to supply truthful propositions is far from certain. However, this should not give any reason to favor

---

<sup>17</sup> by Laura Beth Nielsen, “Power in Public: Reactions, Responses, and Resistance to Offensive Public Speech,” in *Speech and Harm: Controversies Over Free Speech*, by Ishani Maitra and Mary Kate McGowan (Oxford University Press, n.d.), 23.

government restrictions over counterspeech as a better source of truth. I believe the history of suppressive regimes speaks for itself on this subject. Nevertheless, in light of the empirical data, all we can say about this matter is: *who knows* whether the supply of truth will be greater than it would have been under censorial practices.

The empirical basis on which the marketplace metaphor and the counterspeech claim is based is problematic in multiple aspects. My general concern is that the claimants, be they traditional or contemporary commentators on the subject<sup>18</sup>, suppose that more speech is a good remedy against the harmful consequences of bad speech (in itself or in comparison with restrictions) with little or no empirical proof to back it up. And even though I didn't include the sparse amount of research in favor of this supposition,<sup>19</sup> demonstrating the contestability of this empirical proposition is enough for my purposes since its validity (as well as the validity of counterspeech claim and the related truth defense for free speech) is a function of its empirical reliability and demonstrability. Conceptual difficulties further hinder the reliability and the demonstrability of the metaphor as the researches in social sciences don't have an explicitly specified notion of truth to put to the proof. The fact that the empirical research I've cited is not conducted with the general aim of testing the counterspeech claim or the marketplace of ideas also causes additional difficulties when collating and trying to make sense of the data and its relation to the main claim. My conclusion is that we can at best be *agnostic* about the advantages of simply speaking back over the government restrictions.

---

<sup>18</sup> see. Robert Post, "Racist Speech, Democracy, and the First Amendment," *William & Mary Law Review* 32, no. 2 (February 1, 1991): 267; Eugene Volokh, "Freedom of Speech and Workplace Harassment Comment," *UCLA Law Review* 39, no. 6 (1992 1991): 1791–1872.

<sup>19</sup> There is for example research to suggest that groups perform better than individuals in generating reliable information. This supports the idea that "more speech" can lead to truth more reliably, when we understand more speech as more participants, working together. Still, there are a variety of factors that complicate the matter, such as the criteria of better performance, the type of information generated etc. For a more detailed discussion on this subject see. Daniel E Ho and Frederick Schauer, "Testing the Marketplace of Ideas," *New York University Law Review* 90 (2015) (p. 172-4).

## Counterspeech In Practice

Leaving empirical problems with the alleged advantages of counterspeech I now move on to its problematic practical implications. One unfounded assumption that will be the central concern of this section is that there is an inherent tendency or motivation within the people in favor of speaking out. One optimistic advocate speculates along the same lines:

[I]f only truth is left free to combat error, in an open market-place of ideas, humanity is bound to become more enlightened and better off.<sup>20</sup>

The quote above is also a testament to the overarching issue of prescriptive vagueness that advocates of the unsubstantiated counterspeech approach allow. One assumes that as the markets naturally will gravitate towards truth, so will people gravitate towards speaking it out. But as I will try to demonstrate below the facts could not have been more different. Take as an example the cases of defamation: when the issue at hand is an unprotected speech category such as defamation, and the target of defamation is found to be eligible for compensation for the harms they suffered, the courts are expected to be rigorous in their justification and proportionality analysis of the costs that are to be remunerated, the parties responsible and the means available.<sup>21</sup> In this light, it is striking to see some of the highest legal institutions completely ignoring these vital questions and simply proposing to speak back when protected speech poses substantial harms. But who should answer back? How should they? And in fact, can they? These are the questions, answers to which we expect to see in legal and philosophical works that favor a counterspeech response to bad speech. The eerie silence one finds instead has motivated others to approach these questions with higher scrutiny. I will be

---

<sup>20</sup> Bay, Christian. Access to political knowledge as a human right. *Human Context* 7, (1975): 391. As quoted in Howard, "Terror, Hate and the Demands of Counter-Speech," 4.

<sup>21</sup> For a detailed discussion about distribution of harms in cases of protected and defamatory speech see. Frederick Schauer, "Uncoupling Free Speech," *Columbia Law Review* 92, no. 6 (1992): 1321–57, <https://doi.org/10.2307/1122997>.

outlining some of these approaches below, concluding that a just and effective counterspeech remedy requires more explanatory work and more societal and institutional commitment.

The first problem overlooked by the unsubstantiated counterspeech is the fact that the responsibility to answer the hate speakers falls onto the victims of counterspeech. Schauer alludes to a complementary aspect of this issue:

“Still, I will argue that the existing understandings of the First Amendment are based on the assumption that, because a price must be paid for free speech, it must be the victims of harmful speech who are to pay it ... And when in some situations those who bear the cost are those who are least able to afford it, there is even greater cause for concern.”<sup>22</sup>

Schauer’s general point is that free speech principles and constitutional articles protect harmful speech despite the harm it causes to its victims, while society in general benefits from the individual liberty and the security of its crucial dynamics brought about by limiting restrictions. The idea that speech should be protected not because it is harmless but despite its harms is prominent in the literature regarding the regulation of speech, but I’m not trying to pick a fight with that. My concern is the fact that protections of hate speech leave victims in a rare situation where the costs of the common goods fall *disproportionately* on them.

Additionally, the fact that the victims are already socially disadvantaged further confounds the problem. Imagine, having a factory that produces goods that are of significant societal benefit, at the cost of slowly and inevitably making its workers, who don’t have other employment options, sick. Keeping the factory running, and telling the workers that “life is unfair, it just so happens that you folks need to take one for the team” seems not only disingenuous but also deeply, morally troubling. However, the law’s treatment of some hate speech cases resembles exactly that. This becomes a bare-bones counterspeech analogy if we suppose that the factory management not only distributes the whole cost of the social good to the workers but also refuses to pay for healthcare and offers them “the solution” of each worker paying for an

---

<sup>22</sup> Schauer, 1322.

individual plan from their salary. I believe that “more speech” remedy voiced by Justice Brandeis is unfair in a similar way.<sup>23</sup> The fairness objection is intended to supplement the goals of counterspeech that I’ve outlined in the first section. It should guide us when answering the question: “who should answer back?” but by itself, it can’t substitute consequentialist goals of counterspeech as the determinants of successful counterspeech. It should be a matter of fairness on whom the obligation to respond falls, but the harms of hate speech will persist in the absence of an *effective* response.

### **Practical Concerns About Unsubstantiated Counterspeech as a Speech Act**

Philosophers of language have taken issue with unsubstantiated counterspeech arguing, in congruence with what I outlined above, that victims of hate speech often have to fight an uphill battle when answering back.<sup>24</sup> This is connected to but still distinct enough from the pro-regulation line of arguments that claims hate speech and pornography subordinates and silences the targets of such speech. Traditionally, both arguments draw from J.L. Austin’s speech act theory which claims that words can have effects akin to actions and ordinary speakers can have authority through illocutionary force of their speech – given that they fulfill what Austin called “felicity conditions.”<sup>25</sup> Whether the producers of pornographic films, their consumers, or average hate speakers command a level of linguistic authority on

---

<sup>23</sup> I must note that there are significant concerns about the locus of responsibility concerning personal persuasion cases where a speaker (S) convinces an actor (A) to commit a morally impermissible action against a victim (V). Philosophers like Scanlon believe that S can’t be held responsible for the harm V suffers as a consequence of A’s actions. I’m not sure if this scheme and its implications translate well into settings where there is a moral tension between disadvantaged groups and rest of society concerning distribution of benefits and the price to be paid for those benefits. In the factory setting above, I’m not suggesting that factory owners, the government or the rest of the society has to be held directly responsible for the suffering of the workers. My point is that, if there are inevitable costs then there are strong reasons for diffusing the costs and reallocating them towards those who are best able to bear them, since everyone benefits from the goods produced in the factory and people who are currently taking the brunt are already vulnerable enough (they don’t have the option to settle or work somewhere else).

<sup>24</sup> Ishani Maitra and Mary Kate McGowan, eds., *Speech and Harm: Controversies Over Free Speech* (Oxford, U.K: Oxford University Press, 2012); McGowan, “Responding to Harmful Speech”; Rae Langton, “Speech Acts and Unspeakable Acts,” *Philosophy & Public Affairs* 22, no. 4 (1993): 293–330.

<sup>25</sup> John Langshaw Austin and John L. Austin, *How to Do Things with Words* (Harvard University Press, 1975).

par with, say, lawmakers or ship captains is a controversial subject. But the counterspeech counterparts of this argument do not depend on such a strong claim. Mary Kate McGowan argues that a counterspeech strategy that consists of simply challenging hate speech may in fact hamper their counter-speakers' attempts to combat bad speech and cause effects that go against their intentions.<sup>26</sup> According to McGowan one of the many things that we can do with words is to establish new conversational standards and norms. This is the case when the moderator of a video conference (who we assume commands some form of a local authority in the context of the call) says "Participants must write their questions in the chat." With this speech act, a new norm is established where using the chat box is established as a conversationally appropriate thing to do.

However, an explicit invocation of a new rule is not the only way to establish a conversational norm. For example, saying that "Joe (who was born sightless) is the most avid reader of the book club" is not conversationally appropriate. This changes when someone introduces the possibility that the book club is concerned primarily with Braille books. As a more relevant example, a politician publicly stating that "Jewish people secretly control Western financial institutions" makes the question of whether these people in fact command such powers, conversationally relevant. Many trivial and senseless hypotheses like the one above can be easily made relevant to public debate and their harm and dangers are, more often than not, insignificant. The challenge to counterspeech theory lies in the fact that there is an asymmetry between the ease of establishing conversational relevance of a given statement and the challenge of trying to reverse it. Answering the above claim about Jewish people with the simple challenge "Obviously they don't!" or "There is no evidence found to support the claim that they do" makes it *more conversationally appropriate* to discuss whether Jewish people

---

<sup>26</sup> Mary Kate McGowan, "Oppressive Speech," *Australasian Journal of Philosophy* 87, no. 3 (September 1, 2009): 402–4.



control the operation of Western financial institutions. Here I'm not trying to draw attention to stubborn and rationality defying nature of hateful conspiracy theories. My point is that straightforward challenges to substantive hate speech, strengthen the harmful conversational norm that it is okay to discuss, say, whether the Jewish people are nefarious global manipulators.<sup>27</sup> This leaves us with the uncomfortable asymmetry in which conversational norms that are brought about by hate speech easier to establish but harder to reverse by straightforward challenges (which are part of what I've been calling unsubstantiated counterspeech).

The problem is confounded when both the hate and the counter-speaker command a significant level of authority. When the person talking about the influence of the Jews is the Finance Minister, and the straightforward reply comes from the head of the main opposition party, the question becomes more conversationally salient for the public. This effect has been frequently utilized by partisan political campaigns. As Maxime Lepoutre points out, Brexit-Remain campaign's public refutation of the claim that Britain sends £350 million every week to EU ran the significant risk of drawing more attention to this false claim and making it conversationally more relevant.<sup>28</sup> In sum, the asymmetry between the ease of establishing conversational norms and the difficulty of reversing them makes it harder for counterspeech to succeed and its sensitivity to speakers' authority adds to this difficulty. If speech can be rendered ineffective or even counterproductive through certain linguistic practices then by claiming that simply answering back is the solution, the victims of harmful speech are in fact advised to seek a remedy that may prove to be inert against hate speech. These difficulties also haunt arguments that obligate the state to respond on behalf of the victims of hate speech.

---

<sup>27</sup> Admittedly, bringing a matter to public's attention does not constitute harm in itself. But recalling Waldron's point on our strong interest of being perceived as equal citizens, it's fair to say that harm of strengthening discriminatory conversational norms lies in the unwarranted doubt cast on people's good standing as equal and decent citizens.

<sup>28</sup> Maxime Charles Lepoutre, "Can 'More Speech' Counter Ignorant Speech? Tackling the Stickiness of Verbal Ignorance," *Journal of Ethics and Social Philosophy* 16, no. 3 (October 7, 2019): 164.

Even though the range of state counterspeech is not limited to verbal and written statements I will argue that repudiations of discriminatory utterances by the state officials must overcome the linguistic concerns above in order to fulfill the state's obligation of *effective* counterspeech.

### **Epistemic Injustice in the Context of Counterspeech**

Another way that speech can be rendered inert is if the audience has a reduced assumption of credibility towards a certain group, where this assumption is based on epistemically arbitrary reasons. Take the words of Patricia Hill Collins on the subject of images attributed to African American women in the U.S:

Within U.S. culture, racist and sexist ideologies permeate the social structure to such a degree that they become hegemonic, namely, seen as natural, normal, and inevitable. In this context, certain assumed qualities that are attached to Black women are used to justify oppression. From the mammies, jezebels, and breeder women of slavery to the smiling Aunt Jemimas on pancake mix boxes, ubiquitous Black prostitutes, and ever-present welfare mothers of contemporary popular culture, negative stereotypes applied to African American women have been fundamental to Black women's oppression.<sup>29</sup>

Though Collins' work investigates the ramifications of racial and sexist stereotyping across a broad range of contexts, with regards to counterspeech specifically, we can say that the victims of racist and misogynistic speech often have to answer to the offenders (who are likely to hold similar stereotypical beliefs) or (re)establish their standing in the eyes of the public in a state of reduced credibility. Collins counts this effect among the reasons that held back the intellectual tradition of her community for long years. Effects of reduced credibility are quite obvious in certain instances of counterspeech against certain instances of hate speech. In, philosophically speaking, easy cases where we are dealing with immediate and personal instances of counterspeech against hate speech which amounts to little more than fighting words, it's perfectly possible that the hate speaker holds negative opinions as to the

---

<sup>29</sup> Patricia Hill Collins, *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, 2nd ed., Perspectives on Gender (Routledge, 1999), 5.

victim's intellectual capacities (among other bigoted ideas) and that engaging in counterspeech becomes more challenging, if not futile, for the victim. On the other side of the spectrum, the cases involving reputable and authoritative victims of hate speech publicly denouncing impersonal discriminatory ideas (think of Dr. Martin Luther King's *Mountaintop Speech* or his *Letter from a Birmingham Jail*) are not that clear cut.<sup>30</sup> And in between those two examples, there can be any number of hard cases where it's hard to speak of a crystal clear case of reduced credibility. A variety of factors may impinge on counter-speaker's public image and the epistemic authority they possess vis-à-vis their audience and the social sciences face significant challenges in unfurling those factors since most people either incapable or unwilling to easily admit their discriminatory biases. The salience of such biases throughout history and throughout different cultures also may vary wildly. All of these reveal that the effects of reduced credibility on the effectiveness of counterspeech are strongly conditional on a plethora of social and personal contingencies. With all these challenges in mind, it would still be hard to call a counterspeech approach robust or even complete if it didn't account for the effects of reduced credibility on the side of hate speech victims such as racial and sexual minorities. This is an area where empirical sociological research as well as philosophical effort in epistemology needs to progress further before we can demonstrate and contextualize the exact effects of implications of such encounters.

On the bright side, there are noteworthy efforts in epistemology towards this direction. Miranda Fricker argues that epistemic injustice can be the cause of considerable harm for certain groups in society. Related to the subject of counterspeech is testimonial injustice. This boils down to an offsetting influence of speakers' prejudices on their credibility judgments

---

<sup>30</sup> Though today we see those instances, as epitomes of successful counterspeech, we should also be mindful of how many failed attempts of counterspeech it took for the Civil Rights Movement and African American Political activists to get where they were in mid-1960's. Successful and influential examples of counterspeech are not necessarily representative of the general situation.

about others that causes persistent and systematic adverse effects. When a listener holds stereotypical ideas such as “Middle eastern men are hotheaded,” “Orientals are scheming” or “gay couples can’t comprehend the value of traditional family,” these prejudices may weigh in just enough to decrease the credibility afforded to the targets of such ideas. Additionally, we can’t really speak of testimonial injustice in cases of localized prejudices such as a backgammon player being perceived as strategically inept in a meeting of chess grandmasters where the effect is neither persistent (frequently encountered from the target’s point of view) nor systematic (persisting through other dimensions of his life such as economic, professional, educational, etc.). The prevalence of discriminatory stereotypes within different societies and their effect on credibility judgments of different individuals will vary enough to discourage one making sweeping statements about testimonial injustice but we can reasonably assume these two things: Individuals who are engaging in one form of hate speech or other are more likely to hold such prejudices and the reasonable apprehension of a socially or institutionally prevalent prejudice still may have a morally relevant chilling effect on counter-speakers.

Among the examples that Fricker uses to illustrate this effect, she tells the story of a female office worker, who had persistent trouble with getting her message through in the meetings and had to come up with the solution of talking to male colleagues beforehand so they can convey her ideas.<sup>31</sup> The woman said that she doesn’t mind the demeaning and self-sabotaging nature of this workaround as long as the job gets done, but the detriments to her career were obvious: not only she also had to put up with the comments of her supervisors who said she should consider herself lucky to be among a team which constantly comes up with great ideas but also she couldn’t advance in her job because she didn’t receive any credit for her work.<sup>32</sup> More relevant to the subject at hand, Fricker recounts the story of a female

---

<sup>31</sup> Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, 1st ed. (Oxford University Press, USA, 2007), 48.

<sup>32</sup> *Ibid.*, at 47.

Chicana academic<sup>33</sup>, who, because of the false accusations of a white male teaching assistant, had her career undermined. But it was only two years later when a senior male professor concluded that the accusations were baseless, that she started regaining some credibility among her peers. For two years the woman lived with the stress and uncertainty caused by being unable to convince her peers otherwise. Philosophers of language refer to this effect as testimonial incompetence, where the speakers fail to establish credibility or prove their intellectual capacity. For example, Dotson, who employs the same term in her discussion, argues the examples above can constitute a form of chilling effect because the agents may remain silent or irresponsive to abuse when they know it is likely that they won't be understood, taken seriously or perceived as oversensitive or excessively sanctimonious.<sup>34</sup> In this light, expecting from the victims of hate speech, (who may be perceived as intellectually incapable by the speakers or even by a significant portion of society) to answer back and persuade others to the contrary would be optimistic, to say the least. Unsubstantiated counterspeech assumes high responder credibility with no sensitivity to the advantages or disadvantages that a person's social power may bring. In other words, it assumes a *level playing field* between the hate speaker and the victim. The alleged evenness of this playing field has attracted strong criticisms throughout the years, including criticisms concerning counterspeech, and I would like to turn to these now.

---

<sup>33</sup> Originally in Linda Martín Alcoff, "On Judging Epistemic Credibility: Is Social Identity Relevant?," in *Women of Color and Philosophy: A Critical Reader*, ed. Naomi Zack (Oxford: Blackwell, 2000), 248.

<sup>34</sup> Kristie Dotson, "Tracking Epistemic Violence, Tracking Practices of Silencing," *Hypatia* 26, no. 2 (2011): 236–57.

## A Level Playing Field

Responding to a separate discussion on affirmative action, Stanley Fish inveighs against the assumption that there is already a level playing field on which disadvantaged members of the society are free to compete with the better off:

[opponents of affirmative action claim that] “It is undemocratic to give one class of citizens advantages at the expense of other citizens; the truly democratic way is to have a level playing field to which everyone has access and where everyone has a fair and equal chance to succeed on the basis of his or her merit.” Fine words, but they conceal the true facts of the situation as it has been given to us by history: the playing field is already tilted in favor of those by whom and for whom it was constructed in the first place; if the requirements for entry are tailored to the cultural experiences of the mainstream majority, if the skills that make for success are nurtured by institutions and cultural practices from which the disadvantaged minority has been systematically excluded, if the language and ways of comporting oneself that identify a player as “one of us” are alien to the lives minorities are forced to live, then words like “fair” and “equal” are cruel jokes, for what they promote and celebrate is an institutionalized unfairness and a perpetuated inequality.<sup>35</sup>

Implicit here is the idea that liberty from restrictions is enough to create a level playing field. Fish’s point is that not only this is not so, but the pretense of equality hides the influences that shape the plane in favor of one side. I believe there are analogs to be drawn between this and the marketplace metaphor on whose collective audit and good faith the counterspeech claim depends. As Bniku Parekh notes elsewhere, the market may have its own implicit biases and may operate “against the background of prevailing prejudices.”<sup>36</sup> Certain beliefs integral to society may appear common sense, universal or part of what rationality demands. This presents an unfair advantage for these beliefs even before they start to compete with others. This does not have to preclude the chance of the market reflecting on and correcting itself – indeed, what I’m writing here and what many others wrote before me, could be taken as a concrete exercise of this exact possibility. My point is that the market does not come as a

---

<sup>35</sup> Stanley Eugene Fish, *There’s No Such Thing as Free Speech, and It’s a Good Thing, Too* (New York: Oxford Univ. Press, 1994), 62.

<sup>36</sup> Bhikhu Parekh, “Is There a Case for Banning Hate Speech?,” in *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, ed. Michael Herz and Peter Molnar (Cambridge: Cambridge University Press, 2012), 48, <https://doi.org/10.1017/CBO9781139042871.006>.

neutral mechanism by default and at the very least we must vary of this. Moreover, the power of one's speech often depends on the means available to the speaker but protecting the free debate from restrictions is not to say that everybody can have access to powerful means of communication. Imbalance of means and power is something that Mill recognizes within his discussion of the tyranny of the majority.<sup>37</sup> These means could be technological, social, or related to one's unique personality. In this spirit, McGowan argues that the speaker's social position has an impact on whether –or to which extent– she has an audience, on her perceived credibility, and on her perceived expertise.<sup>38</sup> Assuming the existence of a level playing field and saying that one's true propositions have an equal and fair chance to compete against anyone's in the marketplace, is to assume away all these factors which influence one's ability to spread his or her views.

### **Personal Reasons Against Engaging in Unsubstantiated Counterspeech**

With some of the salient moral concerns from the philosophy of language and epistemology outlined above, I would like to move on to a practical criticism of unsubstantiated counterspeech that draws its conclusions from social sciences. I briefly mentioned Laura Beth Nielsen's original research which consists of an analysis of interviews with hate speech and sexual assault victims. The study investigates the personal reasons why the victims of hate speech and verbal sexual assault hold themselves back while ostensibly having an opportunity to speak up. The victims who judge the assailants as ignorant or unreasonably combative may think it is pointless to debate the wrongfulness of the racist remark or may think that there is no use in trying to educate the speaker; they may want to avoid the possibility of getting riled up and answering back in the same, disrespectful manner; or in the case of unsolicited sexual advances by men they may feel a crushing feeling of

---

<sup>37</sup> Mill, *Essays on Politics and Society*, 219–21.

<sup>38</sup> McGowan, "Responding to Harmful Speech," 185.

embarrassment. But Nielsen suggests that the overwhelming reason behind the participants' reluctance to speak back is the concern for their personal safety.<sup>39</sup> The number of people responding to racial slurs and remarks, out of the 100 people participating in the study was less than 5%. The number is less when you deduct the white male subjects who generally have more confidence in their physical safety and have different motivations to respond to racist speech.<sup>40</sup> The persistent fear for one's safety should be read in light of the fact that testimonies of the victims reveal that minority groups and women feel more threatened when they are attacked by the people who they perceive as the socially dominant group.

Nielsen also claims that robust government enforcement of specific policies does carry weight and has transformative potential against the pernicious practices in society. Surely, people may get *some* idea that the practice is inappropriate when the government takes a stance against it. However, it is unclear whether they will be motivated enough by this to overcome the personal aversions against speaking back, and Nielsen does not make clear how and why exactly people will be motivated. Perhaps, the aim should not solely consist of the encouragement of the targets to speak back. This would still leave the unjust burden of speaking back against hate speech on the victim's shoulders and it is not certain that encouragement will skew the linguistic, epistemic, racial, and sex-related power relations between the victim and the assailant enough to allow them to speak back. Perhaps the people, who are supposed to answer back to their assailants, require empowerment and representation. The latter possibility will be the focus of the next section where I outline and discuss Corey Brettschneider's notion of state counterspeech.

In this section, I intended to make clear that targets of hate speech are disproportionately burdened and are not best suited to answer to their assailants. The benefits

---

<sup>39</sup> Ibid., at "Power in Public: Reactions, Responses, and Resistance to Offensive Public Speech," 155–66.

<sup>40</sup> Ibid., at 160.



of speech protections –which everyone enjoys as a social good– leave a small and vulnerable section of people paying for the costs of these benefits who *pro bono* endure the harms of hate speech. But more importantly, I tried to show that people who are supposed to answer back to their assailants often are the ones that lack the linguistic authority, epistemic credibility, self-confidence, and/or feeling of safety to do that. The empirical work so far should prove that the burden of counterspeech, given that it is left on the shoulders of speakers, and given that there are clear power imbalances between them and their assailants, is an *unfair* burden and one that they are unequipped to deal with. Nielsen’s idea that the state has to step in to alleviate some of the harms suffered by the victims is a sensible one but it’s a direction marker at best. With that aim, I think this is a good opportunity to leave behind the criticisms of unsubstantiated counterspeech and move on to a theory of an enhanced, state counterspeech.

## Counterspeech Enhanced

Brettschneider anchors his theory of state expression to the values of freedom and equality which he takes to be crucial for citizens to form political beliefs. For him, theories on the liberal extreme of the free speech debate are problematic because they allow speech that puts values freedom and equality at risk. In contrast, voices advocating for the necessity of regulation and censorship of hate speech are problematic due to the sacrifices that the citizens have to make with regards to their ability to freely and independently cultivate their political beliefs. Ultimately, he aims to develop a third approach to the subject where state abstains from using its *coercive* capacity to curtail speech (thus respecting to the highest degree, the citizen interest in matters related to freedom and equality) and instead uses its *expressive* capacity to influence and persuade the hate speakers. The government action is supposed to be expressive in two senses: expressive in the sense that it protects speech and allows even hate speakers to express themselves, and in the sense that the government expresses its own stance on certain matters with the aim of persuading its citizens.

### Initial Limitations on State Counterspeech

In relation to the expressing its own stance, the state has a myriad of tools at its disposal: written and verbal actions such as anti-discriminatory public statements; symbolic actions such as the removal of artwork and statues belonging to racist historic personalities; subtle and systematic methods such as diverse employment and immigration policies that create a tangible sense of cultural diversity; and economic methods such as financial sanctions against overtly discriminatory enterprises. Though Brettschneider seems to omit this, the government must also comply with its general obligations of accountability and transparency since

clandestine state efforts to manipulate citizens' beliefs would go against everything citizens hold dear in the name of personal autonomy.<sup>41</sup>

Usually what the government can say is also limited by the doctrine of viewpoint neutrality (which is a check on the government not to privilege a particular view over the others), Brettschneider argues that since values of freedom and equality precede and ground the doctrine of viewpoint neutrality, from which the state derives its legitimacy and democratic credentials, the government is permitted to champion these values since in doing so it is not simply expressing a partisan position, it is asserting and explaining the grounds for its legitimacy.<sup>42</sup> The individual nodes and links in Brettschneider's network of values, doctrines, and state legitimacy demand further clarification to reveal the full extent, capability, and shortcomings of his theory and that will be my aim in this section.

The bedrock of Brettschneider's argument is the two-part premise of political liberalism that, the state can't justifiably exercise coercion if it doesn't treat its citizens as free and equal. With a commitment like this, one has to deal with the risk of deeply inegalitarian, authoritarian, theocratic, or apartheid movements coming to power, all of which seek to erode the idea of free and equal citizenship or outright wish to deny it to a certain part of the population. If the premise above suggests that violation of viewpoint neutrality detracts from or annuls the state's legitimacy (which can be understood in terms of its license to use coercion), the government can't take the other side and still expect to be perceived as a legitimate power over its citizens as well as to fulfill its commitments to democracy. Brettschneider sees the first step out of this dilemma in re-defining the groundwork for the

---

<sup>41</sup> The state's expressive range of actions is not completely limited to appealing to citizens' rational capacities. Covert, hypnotic television broadcasts can be ruled out partly because of the transparency concerns I mentioned above, and partly because of the invasive and autonomy denying nature of this method. Certain forms of government counterspeech that subconsciously or emotionally influence the citizens might still be permissible given that these practices are disclosed and given that they follow the other limits outlined throughout the chapter.

<sup>42</sup> Brettschneider, "When the State Speaks, What Should It Say?," 1006.

viewpoint neutrality in non-neutral terms of a full commitment to free and equal citizenship. He draws from the Rawls' framework and claims that political equality is impaired where citizens' capacity for a sense of justice and good faces the threat of unjust coercion. This particular articulation of political morality is certainly not a neutral idea in the framework of viewpoint neutrality, but it can be interpreted as a defense of the same principle since a society where only a proportion of people is able to freely develop and affirm moral and political ideas would not be a society fully committed to the Rawlsian ideal of political equality.<sup>43</sup> Rawls' own writings seem to support this ideal too.<sup>44</sup> Additional grounds for viewpoint neutrality can be found in traditional democracy defenses of free speech which protect speech in the name of citizens' democratic autonomy. According to that, citizens have a strong audience interest hear out all sides of the political arguments; otherwise, they would not truly be the source of their democratic decisions. In addition to that, Dworkin articulates the speaker interest in this regard:

...a majority decision is not fair unless everyone has had a fair opportunity to express his or her attitudes or opinions or fears or tastes or presuppositions or prejudices or ideals, not just in the hope of influencing others, though that hope is crucially important, but also just to confirm his or her standing as a responsible agent in, rather than a passive victim of, collective action.<sup>45</sup>

In sum, democratic autonomy demands that citizens should be able to speak their own and hear others' political opinions. Viewpoint discrimination is impermissible for the state on these grounds because it would deny them the full use of this ability.

Establishing a non-neutral foundation for viewpoint neutrality via a commitment to freedom and equality is a necessary but not sufficient element for the government to give voice to those values. Two other necessary competent that enable this in Brettschneider's

---

<sup>43</sup> Ibid., at 1007.

<sup>44</sup> John Rawls, *Political Liberalism* (Columbia University Press, 1996), 336.

<sup>45</sup> Ronald Dworkin, "A New Map of Censorship," *Index on Censorship* 35, no. 1 (February 1, 2006): 131, <https://doi.org/10.1080/03064220500532412>.

framework are the principles of promulgation and explanation according to which laws must be widely *publicized* to be legitimate and state must justify its use of coercion by an appeal to *reasons* respectively.<sup>46</sup> This necessitates the state to make sure that the laws are accessible to its citizens, and that citizens know when and why they are subject to coercion. The state may –and according to Brettschneider *should*– make use of this avenue when faced with expressions and political movements that oppose the ideals of freedom and equality: “The legitimate state, acting in its ‘expressive’ capacities as opposed to its coercive ones, then has an obligation to explain why certain viewpoints that are protected from being silenced by certain rights are at odds with those same rights.”<sup>47</sup> This is neither a novel nor a controversial idea in actual, legal contexts. Indeed, when ruling that forms of speech which at odds with is democratic values are protected, constitutional and international courts, as well as their lower courts, give reasons for their decisions and usually make clear that their ruling is not an endorsement<sup>48</sup> of the views in question. Brettschneider thinks institutions such as the U.S. Supreme Court that function this way in the name of free and equal citizenship makes them good models for other government bodies and public officials.<sup>49</sup>

Brettschneider grounds the expressive obligation of state not only in terms of legitimacy but also in terms of an urge not to appear complicit in hateful expressions. He points out that the state’s message for its actions, in addition to being publicized, is often reflected in the actions themselves. So, for instance, when a state regulates the use of religious symbols in its institutions, through his actions it communicates that the state is not officially endorsing any religion. In contrast, when the state refuses to regulate hate speech the message is not that clear, judging by the action alone it can be said that the state takes no side in the

---

<sup>46</sup> Brettschneider, “When the State Speaks, What Should It Say?,” 1008.

<sup>47</sup> Ibid., at 1008.

<sup>48</sup> For a discussion on causal nuances between complicity in expression and endorsement see Paul Billingham, “State Speech as a Response to Hate Speech: Assessing ‘Transformative Liberalism,’” *Ethical Theory and Moral Practice* 22, no. 3 (2019): 639–655, <https://doi.org/10.1007/s10677-019-10001-1>.

<sup>49</sup> Ibid., at “When the State Speaks, What Should It Say?,” 1008.

debate about hate speech, or takes the side of the hate speech. This suggests that in freedom expression cases involving speech that is at odds with values of free and equal citizenship, there is an obligation on the state's side for clarification of its reasons for doing so:

When the state refrains from regulating illiberal viewpoints, it is essential that it also use its expressive capacities to clarify that it is not expressing support for the viewpoints themselves, but instead is guaranteeing an entitlement that stems from the need to respect all citizens as free and equal.<sup>50</sup>

In addition to the clarificatory message to the audience who are the targets of and bystanders in relation to hate speech, state counterspeech speaks to the hate speakers and says that while their speech is protected the content of their expression contradicts the reasons behind the protection of their speech. This points toward one of the primary aims of state counterspeech: influencing and persuading the hate speakers. Brettschneider calls the culmination of such state efforts *democratic persuasion* in which the state offers justifications for its laws or citizens' rights with the aim of the citizenry, especially the hate speakers, accepting and internalizing the reasons behind those justifications. It must be noted that however reasonable and appealing this attempt to persuade the hate speaker might be, it is still up to the hate speakers whether to give credence to the persuasion attempt.

### **Two Further Limits on State Counterspeech**

Brettschneider argues that the state should move towards more effective means of counterspeech but while doing so it must take care that it observes two limitations.<sup>51</sup> First of these dictates that, the method of influence must not be in violation of fundamental rights that citizens hold, such as the freedom of expression, religious belief, and association. According to this view, banning racist rallies on account of their ideology would be impermissible while changing the curriculum to be more sensitive to the subject of discrimination or funding

---

<sup>50</sup> Ibid., at 1010.

<sup>51</sup> Ibid., at Brettschneider, 1010.

human rights groups that oppose racist messages would not.<sup>52</sup> This may suggest that the state appeals to freedom and equality both while justifying its right to speak and when allowing discriminatory speech and action. But note that the pressure on the side of the state to speak out against hate speech is rooted in its commitment to the ideals of free and equal citizenship, while its commitment to uphold its citizens' rights to free speech, religious liberty and freedom of association (all of which can be exercised to undermine freedom and equality) may depend on other principles –indeed this is the case with autonomy or democracy related interests in the context of free speech. In the case of state counterspeech, this certainly creates tension between competing principles, interests and values (as numerous debates between the proponents of different free speech principles can attest) but I don't believe that it leads to a paradoxical situation in which the state appeals to a single set of values to justify the exercise of certain rights that undermine each other.

Brettschneider calls this the “means-based limit” and it circumscribes the limits of government coercion on its subjects, in relation to its attempt to change their views and influence their expressions. Means based limit is an expression of state's commitment to a distinct free speech principle which is (1) more specific than a negative right to liberty where the state abstains from interference with the citizens' freedom of action as much as possible; and which (2) goes beyond a basic harm principle and protects hate speech despite the harms in the name of other free speech interests.

Notice that the definition of coercion will be directly proportional to what extent the government can justifiably raise its voice –so to speak– while engaging in counterspeech. So, it does a lot of work in the argument for state counterspeech and I would like to spend a few moments outlining his understanding of coercion.

---

<sup>52</sup> Government spending and its influence on education policy are by no means uncontested subjects. This will play a role in my criticism of Brettschneider at the end of this section.

Brettschneider defines coercion as “the state threatening an individual or group of individuals with a sanction or a punishment with the aim of prohibiting a particular action, expression, or holding of a belief.”<sup>53</sup> In this view, the outright prohibition of dangerous or harmful political expression would be impermissible as it is detrimental to citizens’ ability to independently form political beliefs. On the other hand, state counterspeech with transformative aims in mind does not qualify as coercion because by doing so “the state does not seek to prohibit [its subjects] from holding conflicting beliefs.”<sup>54</sup> Channeling the spirit of Dworkin, he adds that it is crucial for the ideal of free expression that citizens are given a chance to reject the persuasive attempt by the state. A *meaningful* chance of rejecting government-backed ideas and policies is necessary to preserve citizens’ autonomy-related speech interests and ensure their freedom of conscience.

The question of whether some methods of state counterspeech actually respect this absolute minimum is relevant and I will touch on this subject when discussing some of the methods that state can use, but before that, it would be useful to say a few words on the other limit on government’s range of transformative actions.

The substance-based limit demands the government confine its persuasive effort to the expressions that are at clear odds with values of free and equal citizenship. This rules out inegalitarian beliefs which are not directly hostile to the ideals of free and equal citizenship as targets of democratic persuasion. The issue about where the limits of open hostility exactly lie is unclear. Brettschneider’s definition appears to include groups like Ku Klux Klan while leaving out the opponents of affirmative action or advocates of colorblind hiring practices. The rationale behind that seems to be the benefit of reasonable doubt on the side of the latter since there can be plausible yet controversial interpretations of equality that argue affirmative

---

<sup>53</sup> Brettschneider, “When the State Speaks, What Should It Say?,” 1010.

<sup>54</sup> *Ibid.*, at 1010.



action asks too much of the state or that members of society have access to a sufficiently level playing field.

With the framework, available theoretical means, and limits in place we can finally discuss the actual instances of state counterspeech that Brettschneider endorses. The operating range of democratic persuasion is not limited to offering reasons or even to state criticism of hate groups as this is made clear by the concrete, specific examples he offers as well as his approach to the role of certain state duties such as education and spending. The Senate confirmation hearing of Supreme Court Justice candidate provides an illustrative account of Brettschneider's general scheme. Justice Samuel Alito's confirmation hearing in 2006 in front of the U.S. Senate, among other concerns, brought his membership to a discriminatory student club to public discussion. The club, called Concerned Alumni of Princeton (CAP) advocated against coeducation and enrollment of minorities in the university and itself was reluctant to admit women or minorities of African American origin. Brettschneider argues that the Senate Judiciary Committee's inquiry into and criticism of Alito's membership to the club as well as its (unrealized) potential to veto him down in case he did not denounce the club's message is a matter of permissible democratic persuasion.<sup>55</sup> Alito was indeed criticized and asked about his affiliation to the club which he falsely answered in negative. More importantly, he was insistently asked whether he was "against women and minorities attending colleges" to which he answered "Absolutely not, Senator. No."<sup>56</sup> Supreme court justices are expected to rule on cases involving discrimination and they are expected to adhere to the ideals of an egalitarian and democratic state. This is made clear in the code of conduct for the U.S. judges: "A judge should not hold membership in any organization that practices invidious discrimination on the

---

<sup>55</sup> Ibid., at 1012.

<sup>56</sup> "U.S. Senate Judiciary Committee Hearing on Judge Samuel Alito's Nomination to the Supreme Court," accessed June 13, 2020, <http://www.washingtonpost.com/wp-dyn/content/article/2006/01/10/AR2006011000781.html>.

basis of race, sex, religion, or national origin.”<sup>57</sup> Thus we can say that the committee was obliged to bring this issue into public attention and to openly criticize in a televised and otherwise widely publicized confirmation hearing. Brettschneider rightly points out that through their criticism the senators acted in their capacity as public officials and exemplars with the transformative aim of persuading citizens that CAP’s views have no place in the highest court of the U.S.<sup>58</sup> This was not in violation of the substance based limit as the pernicious views in question were specifically and openly against the values of freedom and equality. More importantly, Brettschneider argues that it would not have been in violation of means-based limit if Alito was vetoed down on the basis of his membership to or endorsement of CAP since the codes of conduct on which the committee based its inquiry and criticism does not outlaw joining a discriminatory club or prohibit the views, they just say such people can’t hold the office of a Supreme Court justice;<sup>59</sup> ergo: the state would have acted in its persuasive capacity as opposed to its coercive capacity. This, in the broad framework of Brettschneider’s understanding of coercion, is a point that attracted recent criticism, and I will be also focusing on this at the end of this section.

But before that let us also look at other tools that are available to state in its expressive role. Possibilities include government championing the values of freedom and equality by instituting public holidays, officials giving statements on record against discriminatory views, production of radio and television programs, commissioning the construction of memorials, and creating specific endowments for arts dealing with the matters of gender or race-based discrimination.<sup>60</sup> All those are compatible with Brettschneider’s limitations as long as the

---

<sup>57</sup> “Code of Conduct for United States Judges,” United States Courts, accessed June 13, 2020, <https://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges>.

<sup>58</sup> Brettschneider, “When the State Speaks, What Should It Say?,” 1012.

<sup>59</sup> Ibid., at 1012. The exact wording of coercion at page 1013 is also important and will be subject to criticism. Brettschneider claims “that in the Alito hearing, the committee members *did not attempt to force* Alito to reject the values of the Princeton club” (my italics).

<sup>60</sup> Similar proposals are also listed in Gelber, “Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia),” 214.

measures are not coercive, the dissenting citizens reserve the right to oppose those measures and the target of state counterspeech is the expression that is clearly hostile to freedom and equality.<sup>61</sup> State influence on educational policy would also be permissible as long as teachers do not resort to repressive methods in the classroom. The state can and should include strong criticisms of discriminatory and eliminationist practices in the official curriculum and actively praise historic figures who stood against them all the while giving the dissenting students a chance to argue their side in civil debate. In keeping with his stance on the confirmation hearing of Justice Alito, Brettschneider argues that states can fire employees who clearly oppose the values that state should endorse, such as teachers who are active Klan members or advocate the policies of KKK in or even outside the classrooms: “the hateful viewpoint of a public worker [should not be] be protected at the expense of the state’s ability to explain the reasons for free speech ... I would even argue that this kind of hateful expression by the teacher could be grounds for dismissal even if done outside the classroom or the school.”<sup>62</sup> That’s another point that I will return to but for now, it is enough to say that Brettschneider seems at least unclear on the amount of coercion involved in democratic persuasion.

Last and arguably the most effective expressive act that Brettschneider allows government is to shape its spending policy in light of its obligations to freedom and equality. He tries a prophylactic move against the criticism that the state restricting or cutting funding against a certain group crosses the line between persuasion and coercion. The idea is that by cutting funding or canceling the tax-exempt status of a hate group, the state does not *violate* but rather *diminishes* their right to free expression which was protected by what he calls the means-based limit within his framework. According to that, a total violation would consist of

---

<sup>61</sup> I would argue that, governments tearing down the statues dedicated to overtly racist and totalitarian personas would also pass Brettschneider’s coercion and appropriate content test though he doesn’t mention this specific act of state expression.

<sup>62</sup> Brettschneider, “When the State Speaks, What Should It Say?,” 1013.

a world where the state funding or the tax-exempt status was the only available income that facilitated hate groups expression. Since those groups have other means to fund their expressive activities, Brettschneider claims, this does not constitute a violation of their free speech rights. Of course, diminished rights can simply be said to be violated to a lesser extent—an obvious linguistic fact that Brettschneider fails to consider. Add to that the fact that free speech rights are usually taken as negative rights. That is, they circumscribe the limits of government interference with citizens' free speech while saying nothing about the obligation on the government's side to provide the means for speech. Adverse changes in state spending and taxation policy indeed diminish the speakers' positive rights to speech but Brettschneider doesn't argue that citizens have such claims and it is doubtful whether they in fact have a just claim for a positive right to free speech. Since his theory neither defines a distinction between diminished and partially violated rights nor does it offer any substantive comment about the negative nature of free speech rights, I take this argument to be incomplete at best and contradictory vis-à-vis limits described in author's own framework at worst. Furthermore, Brettschneider argues that financial sanctions that he proposes do not constitute coercion by his definition:

offering financial inducements, like pure persuasion, is clearly an attempt to convince citizens to make a particular choice, *but it does not deny the citizen the right to reject that choice*. While potential state funding can serve as an incentive to believe certain ideas, citizens might legitimately choose to reject those beliefs and forgo those funds.<sup>63</sup>

This comes right after he agrees with the claim that the government imposing parking fines in certain zones is coercive. It is coercive because the government intends its citizens not to park in said zones and—somehow in this instance but not in the case of financial sanctions against hate groups—“does not allow citizens to choose whether they wish to pay to park in a red zone.”<sup>64</sup> From where I'm standing it looks like in both cases, I do have a choice to pay the

---

<sup>63</sup> Ibid., at 1015–16 (my italics).

<sup>64</sup> Ibid., at 1015.

price, suffer the consequences and do the action anyway: in this hypothetical scenario, I do have a choice to violate governments parking regulations and bear the costs, and I don't see any possible way how the government can deprive me of that choice. My point is that the financial sanction case and parking violation case are not *obviously* dissimilar and if the latter is coercive, I don't see any explanation in Brettschneider's article why this isn't the case for the former.

Paul Billingham's criticism of Brettschneider may help to clarify this matter. Billingham approaches this topic through a contrast between Brettschneider's definition of coercion and Nozick's, which is in fact the model for Brettschneider's.<sup>65</sup> Remember that Brettschneider's definition was: "the state threatening to impose a sanction or punishment on an individual or group of individuals with the aim of prohibiting a particular action, expression, or holding of a belief"<sup>66</sup>—potential sanction or punishment being the cancellation of tax-exempt status in this case. I take it that the state's aim of threatening financial sanctions is to make certain groups change their minds on policies that are in deep conflict with the values of freedom and equality. For Brettschneider, this does not amount to coercion because the group has the choice of sticking to its beliefs at the expense of potential financial loss. I agree with Billingham in that "[this] allows any threat to be re-interpreted as non-coercive."<sup>67</sup> For instance, in the parking violation case, the government may deny coercion on its side since, as I said above, I have the choice of paying the fee and parking there anyway.<sup>68</sup> As a concrete example on this matter, we can look at Brettschneider's comments about financial state sanctions against two discriminatory organizations: "although their tax subsidies were

---

<sup>65</sup> Billingham, "State Speech as a Response to Hate Speech."

<sup>66</sup> Brettschneider, "When the State Speaks, What Should It Say?," 1010.

<sup>67</sup> Billingham, "State Speech as a Response to Hate Speech," 645.

<sup>68</sup> Scanlon makes a similar point, claiming that speakers can succeed in their exercise of autonomy even if they are punished as a consequence if their expression, but this can't be taken as a scenario where people are free to speak their minds. In Thomas Scanlon, "A Theory of Freedom of Expression," *Philosophy and Public Affairs* 1, no. 2 (1972): 204–226.

discontinued, Bob Jones University and the Christian Legal Society continued to exist and exercise their right to dissent.”<sup>69</sup> That is correct, but not the same thing as claiming that they are not worse off than they could have been. Presumably, the state sanctions have incurred operational costs on these groups and their continuing existence hangs from a thinner thread as a result. On the other hand, they can be claimed not to be coerced only in the sense that the government’s threat *failed to change their minds* and the question of whether they were unjustly punished for their discriminatory beliefs, remains there to disconcert most liberals in support of Brettschneider’s theory.

In contrast, Nozick’s own classical formulation of coercion does not seem to allow the interpretation above. For Nozick, P coerces Q when:

1. P aims to keep Q from choosing to perform action A;
2. P communicates a claim to Q;
3. P’s claim indicates that if Q performs A, then P will bring about some consequence that would make Q’s A-ing less desirable to Q than Q’s not A-ing;
4. P’s claim is credible to Q;
5. Q does not do A;
6. Part of Q’s reason for not doing A is to lessen the likelihood that P will bring about consequence announced in (3)<sup>70</sup>

One can’t follow Brettschneider if he wants to avoid the risk of ending up with an understanding of coercion that allows any threat to be interpreted as non-coercive. According to Nozick’s definition, if (5) obtains, parking violation fines are coercive (which is what the result that we expect from a definition of coercion to offer and what the false interpretation above fails to deliver) but so are certain forms of democratic persuasion such as reduction or cancellation of state funding. This criticism puts pressure on Brettschneider’s means-based limit as some state actions threaten to violate it. Facing this challenge, state counterspeech either has to admit that government’s role as a spender (and possibly as an employer if we

<sup>69</sup> Corey L. Brettschneider, “Democratic Persuasion and Freedom of Speech: A Response to Four Critics and Two Allies,” *SSRN Electronic Journal*, 2014, 1082, <https://doi.org/10.2139/ssrn.2681229>.

<sup>70</sup> As quoted in John McMillan, *The Methods of Bioethics: An Essay in Meta-Bioethics* (Oxford University Press, 2018), 171.

apply a parallel criticism against Brettschneider's justification for dismissal of employees) does not fall within the range of permissibility of democratic persuasion or the proponents of state counterspeech may need to abandon the means-based limit altogether in favor of a framework that admits more coercion than some forms of political liberalism are ready to accept. I believe that the second option is closed to Brettschneider due to his commitment to very a strong form of political liberalism where justification of state restrictions is more theoretically challenging.

An attractive solution to this requires re-examining the institutional claims to tax-exempt status. Note that both Brettschneider and Billingham assume a baseline in which certain societal intuitions are entitled not to pay taxes in exchange for their charitable activities. Their tax-exempt status has to be taken within the conjecture of a variety of political and historical reasons and it is contingent on the nature and effectiveness of the activities that they perform (in an ideal scenario certain NGO's operating in Hungary would not be paying the so-called anti-immigration tax<sup>71</sup>, while the Church of Scientology would be required to pay income and property taxes<sup>72</sup>). In this light, it is possible to see the tax-exempt status as a privilege that the government allows for certain entities and not as a claim on the side of the receiving institutions. With this baseline (which also applies to other forms of government spending) a new approach to coercion can be adopted where the government threatens only to withdraw a privilege instead of taking away certain rights that some institutions hold. If we operate in a reference frame where no one has an initial claim to tax-

---

<sup>71</sup> "Hungary's Anti-NGO Tax Violates Free Speech and Freedom of Association," accessed August 26, 2020, <https://www.justiceinitiative.org/voices/hungary-s-anti-ngo-tax-law-violates-free-speech-and-freedom-association>.

<sup>72</sup> "What Is the Significance of the IRS Ruling Regarding Churches of Scientology?," Official Church of Scientology: What is Scientology?, accessed August 26, 2020, <https://www.scientology.org/faq/church-funding/significance-of-irs-ruling.html>.

exempt status, then losing that status after a while does not make one worse-off compared to their baseline.

On the other hand, lesser commitments to political liberalism, such as free speech justifications which allow for the regulation of some speech, might endorse state counterspeech instead of or in addition to restrictions with little threat from criticisms above.

The argument from state complicity in its citizen's harmful and dangerous beliefs also deserves some scrutiny. I take this argument's function within the whole of Brettschneider's theory to be secondary, therefore if found suspect and abandoned it would not leave remaining parts of his theory more vulnerable than they were before. As a reminder, Brettschneider argues that it is essential for the state to "use its expressive capacities to clarify that it is not expressing support for the viewpoints themselves."<sup>73</sup> The first problem with this is the alleged need to clarify: Brettschneider does not offer any empirical support for the claim that citizens may perceive the government as complicit in hateful expressions. An additional question is whether the alleged perception of the state is serious or widespread enough to justify all the requirements of his framework of state counterspeech. Again, we have don't have sufficient, reliable information to answer this question. Even according to the most charitable reading of this claim, where a widespread perception of state complicity exists, the state can point out that if it is complicit in hateful expressions, it should be complicit in all the other expressions that it allows and protects.<sup>74</sup> Does the state have to issue clarifications on its protection of flat-earther theories? What about bot activity on Twitter? At the very least, the argument from complicity seems to be in conflict with the substance-based limit. All in all, it is more reasonable and less theoretically demanding for the state to make the argument against the equivocation of protection to endorsement instead of resorting to expressing his

---

<sup>73</sup> Brettschneider, "When the State Speaks, What Should It Say?," 1010.

<sup>74</sup> Billingham, "State Speech as a Response to Hate Speech," 643.



view through funding cuts. Note that the obligation of the state to clarify and explain its reasons for its law remains intact per principles of promulgation and explanation – counterparts to which can usually be found in various legal systems. Therefore, I argue that argument from state complicity does more harm than good to the overall theory of state counterspeech and can be abandoned without much danger if one wishes to maintain his or her commitment to unproblematic parts of the theory.

Lastly, I would like to draw attention to two aspects of Brettschneider's ideal theory that can negatively affect its application in real life. These are the issues of overbreadth and state distrust, both stemming from Brettschneider's commitment to the substance-based limit and a particular understanding of free and equal citizenry. According to this commitment, only the groups and individuals "hostile to or implausibly compatible with"<sup>75</sup> the values of free and equal citizenship can be legitimate targets of democratic persuasion. According to this definition, KKK fits the bill perfectly: historically, the Klan unashamedly and openly advocated for the exclusion of African American minorities from the white American society, and its arguments that invoke equality in specific contexts (such as restoring equality for the white race) can't be taken as plausibly compatible with the ideal of free and equal citizenship. Meanwhile, political positions who reasonably disagree with some claims of egalitarianism are ruled out as targets of democratic persuasion. As is often the case, this standard of reasonableness is a blurry line and may grant a lot of discretion on the side of the adjudicator. This worry invades the non-ideal plain of the theory because of Brettschneider's comments about the specific cases. As I mentioned above, he is mostly on board with the constitutional decisions against Bob Jones University and Christian Legal Society. Both organizations had gender or race-based discriminatory admission policies for which they lost or were denied their claim to special status. Defenders of these organizations have made the highly dubious

---

<sup>75</sup> Ibid., at "When the State Speaks, What Should It Say?," 1011.

and disingenuous claim that excluding someone from an association does not say anything about that person's worth or status within the society. Still, on the most charitable reading, I guess these can be taken as *plausible* arguments that invite further discussion. On the other hand, responding to a comment on the Roman Catholic Church's policy of not admitting women to priesthood Brettschneider says this: "The Church, however, has taken the position that restricting the priesthood to men does not imply that it opposes the equal status of women" and "when the issue is ambiguous we should give the organization the benefit of the doubt about the consistency of its position with the ideal of free and equal citizenship."<sup>76</sup> The Brettschneider leniency and inconsistency toward the Roman Catholic Church is not the main problem, nor does this fact defeat his general theory in one swift stroke. When his critics have questioned his stance on a specific subject<sup>77</sup>, he responded<sup>78</sup>, in turn, others pointed out additional problems and inconsistencies<sup>79</sup>, and we can expect Brettschneider to respond to them too. The point is that the compatibility issues can be extremely complicated and they require adjudication of ambiguous judgments about religious doctrines and internal politics of the organizations. Many would agree that the public debate in a vibrant civil society is better equipped than a state committee to adjudicate the matter of compatibility of policies with the ideals of free and equal citizenship. On the subject of state distrust, Paul Billingham comments that:

Liberals should particularly fear that the state is likely to intervene too much, demanding unreasonable levels of congruence, in ways that excessively interfere with the freedom of speech and association of groups whose views or practices do not fully cohere with liberal egalitarian ideals.<sup>80</sup>

---

<sup>76</sup> Brettschneider, "Democratic Persuasion and Freedom of Speech," 1084.

<sup>77</sup> Steven G Calabresi, "Freedom of Expression and the Golden Mean," *BROOKLYN LAW REVIEW* 79 (2014): 11; Sarah Song, "The Liberal Tightrope: Brettschneider on Free Speech," *BROOKLYN LAW REVIEW* 79 (2014): 12.

<sup>78</sup> Brettschneider, "Democratic Persuasion and Freedom of Speech."

<sup>79</sup> Billingham, "State Speech as a Response to Hate Speech."

<sup>80</sup> *Ibid.*, at 648.

The risk is to bestow a dangerous amount of initial discretionary power to the state, and the fact that we have the option to dissent doesn't recuperate the sunk costs. These costs are not only related to costs that organizations hostile to the idea of free and equal citizenry will have to endure; there also maybe also a societal cost in terms of *a chilling effect on the civil society*. The trouble here is the potential overbreadth caused by an extensive understanding of free and equal citizenship ideal. Remember that certain people were liable to be fired from their jobs according to Brettschneider's theory, even if they did not express their beliefs against free and equal citizenry in their capacity as government employees. In Brettschneider's writings, there is evidence to suggest that the ideal free and equal citizenry is incompatible with conservative positions such as advocacy of traditional family roles and opposing gay marriage.<sup>81</sup> That means, conservatives who make public statements in support of these positions while reluctantly treating their co-workers and pupils in an egalitarian fashion are legitimate targets of democratic persuasion and are liable to lose their jobs. Billingham somehow optimistically but plausibly claims that: "many such individuals are good-willed, respectful, and civil. They do not think others are of less moral worth. They simply hold traditional moral views that many consider incorrect."<sup>82</sup> Many charitable religious groups will also be liable to lose their tax-exempt status or government funding, causing or contributing them to shut down their operations. Given the state's possible margin of error in adjudicating complex cases and the potential overbreadth issues regarding the incompatibility of certain positions with the ideals of the free and equal citizenry, the cost of misguided democratic persuasion can be a less vibrant and less effective civil society. I completely agree with Billingham in that there is substantial value in a healthy civil society, and this requires:

accepting that some groups we do not like, or consider deeply mistaken or even positively harmful, will exist and enjoy the same public status—including tax

---

<sup>81</sup> Brettschneider, "When the State Speaks, What Should It Say?," 1011; Brettschneider, "Democratic Persuasion and Freedom of Speech," 1084.

<sup>82</sup> Billingham, "State Speech as a Response to Hate Speech," 651.

exemptions—as groups that we favour. This is a price worth paying for the sake of the broader goal of securing a diverse and active associational sphere, with all of the benefits that this brings.<sup>83</sup>

All in all, I believe that criticisms above neither extinguish the need for a better form of counterspeech nor rule out the possibility of state involvement on the side of the speaker in hate speech cases. First of all, one doesn't have to subscribe to the entirety of traditional liberal thought, to accept and work with the criticism above with the aim of bolstering and modifying Brettschneider's theory. As I mentioned at the outset, counterspeech is a notion that is compatible with multiple views on state regulation of speech. I think that free speech justifications that allow for some of the bad speech regulated and allow for more types of state coercion to be justified, can gain a lot from Brettschneider's state obligation for counterspeech. But I also believe that his arguments are available even to traditional liberals given that there is a reduction of scope concerning the targets of democratic persuasion and a narrowing or specification of definition of free and equal citizenry ideal. I must also add that, in the end, Brettschneider's theory is a form of enhanced counterspeech that mostly focuses on transformative goals of counterspeech—that is, on the aim of changing the speakers' minds. The lack of emphasis on the reparative capacity of counterspeech is a concession that he has to make due to his commitment to protecting hate speech. So, while state support definitely bolsters the standing and efficacy of counterspeech compared to unsubstantiated versions of it, a full theory must also take heed to the audience's interests in avoiding the harms of speech.

As it stands Brettschneider's state counterspeech account does not take much notice with the concerns discussed outlined in the previous section. The fact that the state is better equipped to speak back against hate speech compared to the victims should not stop the state

---

<sup>83</sup> Ibid., at 654.

from endeavoring to become more effective in its expressions. The state's obligation to speak back would be tantamount to lip-service if the fulfillment requires *simply speaking back*.

## Conclusion

The deficiencies of the unsubstantiated counterspeech approach led us to look for an enhanced solution to combatting the harm of hate speech. Asking the victims to speak back to their assailants or to wage a solitary crusade against the expressions of systematic discrimination is asking too much of them while their prospects of accomplishing anything this way is dim. Government representation can further reparative and transformative aspirations of counterspeech and the grounds for justification for this intervention exist in legitimate states' commitment to freedom and equality. While exploring this possibility I tried not to let my guard down concerning the subject of unduly empowering the government. State counterspeech, albeit attractive, is still delegating more authority to the state in an area where it had not yet a chance to prove its trustworthiness. Due to the historical track record of governments concerning speech regulations, we must be extra careful while inviting it to the business of convincing others.

Indeed, most of the work done on the subject of free speech concerns speech regulations, which are seen by many as blunt instruments of questionable efficacy and moral permissibility. As I said, I also share some of these concerns. But there is also work being done with the aim of preventing harm to the targets of counterspeech both in the natural sciences and philosophy. Jeffrey Howard<sup>84</sup> for example is equally sensitive to reparative as well as transformative aims of counterspeech in his theory where he suggests a samaritan, civic duty to speak back. Philosophers of language also started paying much closer attention to the mechanics of counterspeech in the last decade and influential philosophers like Rae Langton<sup>85</sup> and Mary Kate McGowan<sup>86</sup> have argued for their own forms of linguistically

---

<sup>84</sup> Howard, "Terror, Hate and the Demands of Counter-Speech."

<sup>85</sup> Langton, "Blocking as Counter-Speech."

<sup>86</sup> McGowan, "Responding to Harmful Speech."

effective counterspeech, in frameworks distinct from the debates surrounding regulation. However, each of these commentators has substantive commitments to widely debated and sometimes conflicting philosophical doctrines and it is a nigh-impossible task to combine those in a grand theory of counterspeech without lengthy discussions about their own particular views on morality and language. It will be sufficient to say that, highly needed and appreciated work is being done on the subject of counterspeech where it is being offered as a distinct, robust solution instead of a moral consolation prize when the regulations are impermissible.

## References

- Abrams v. United States, 250 U.S. 616 (1919).
- Austin, John Langshaw, and John L. Austin. *How to Do Things with Words*. Harvard University Press, 1975.
- Billingham, Paul. "State Speech as a Response to Hate Speech: Assessing 'Transformative Liberalism.'" *Ethical Theory and Moral Practice* 22, no. 3 (2019): 639–655. <https://doi.org/10.1007/s10677-019-10001-1>.
- Brettschneider, Corey. "When the State Speaks, What Should It Say? The Dilemmas of Freedom of Expression and Democratic Persuasion." *Perspectives on Politics* 8, no. 4 (December 2010): 1005–19. <https://doi.org/10.1017/S1537592710003154>.
- Brettschneider, Corey L. "Democratic Persuasion and Freedom of Speech: A Response to Four Critics and Two Allies." *SSRN Electronic Journal*, 2014. <https://doi.org/10.2139/ssrn.2681229>.
- Calabresi, Steven G. "Freedom of Expression and the Golden Mean." *BROOKLYN LAW REVIEW* 79 (2014): 11.
- United States Courts. "Code of Conduct for United States Judges." Accessed June 13, 2020. <https://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges>.
- Collins, Patricia Hill. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd ed. Perspectives on Gender. Routledge, 1999.
- Dotson, Kristie. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26, no. 2 (2011): 236–57.
- Dworkin, Ronald. "A New Map of Censorship." *Index on Censorship* 35, no. 1 (February 1, 2006): 130–33. <https://doi.org/10.1080/03064220500532412>.
- Fish, Stanley Eugene. *There's No Such Thing as Free Speech, and It's a Good Thing, Too*. New York: Oxford Univ. Press, 1994.
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. 1st ed. Oxford University Press, USA, 2007.
- Gee, Gilbert C. "A Multilevel Analysis of the Relationship Between Institutional and Individual Racial Discrimination and Health Status." *American Journal of Public Health* 92, no. 4 (April 2002): 615–23.
- Gelber, Katharine. "Reconceptualizing Counterspeech in Hate Speech Policy (with a Focus on Australia)." In *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, edited by Michael Herz and Peter Molnar, 198–216. Cambridge: Cambridge University Press, 2012. <https://doi.org/10.1017/CBO9781139042871.016>.
- Heath, Chip, Chris Bell, and Emily Sternberg. "Emotional Selection in Memes: The Case of Urban Legends." *Journal of Personality and Social Psychology* 81, no. 6 (2001): 1028–41. <https://doi.org/10.1037/0022-3514.81.6.1028>.
- Ho, Daniel E, and Frederick Schauer. "Testing the Marketplace of Ideas." *New York University Law Review* 90 (2015): 69.
- Howard, Jeffrey W. "Terror, Hate and the Demands of Counter-Speech." *British Journal of Political Science*, 2019, 1–16. <https://doi.org/10.1017/S000712341900053X>.
- "Hungary's Anti-NGO Tax Violates Free Speech and Freedom of Association." Accessed August 26, 2020. <https://www.justiceinitiative.org/voices/hungary-s-anti-ngo-tax-law-violates-free-speech-and-freedom-association>.
- Langton, R. "Blocking as Counter-Speech." In *New Work on Speech Acts*, edited by Daniel Harris, Daniel Fogal, and Matt Moss, 144–64. New York: Oxford University Press, 2017. <https://doi.org/10.1093/oso/9780198738831.003.0006>.
- Langton, Rae. "Speech Acts and Unspeakable Acts." *Philosophy & Public Affairs* 22, no. 4 (1993): 293–330.



- Lepoutre, Maxime Charles. "Can 'More Speech' Counter Ignorant Speech? Tackling the Stickiness of Verbal Ignorance." *Journal of Ethics and Social Philosophy* 16, no. 3 (October 7, 2019). <https://doi.org/10.26556/jesp.v16i3.682>.
- Maitra, Ishani, and Mary Kate McGowan, eds. *Speech and Harm: Controversies Over Free Speech*. Oxford, U.K: Oxford University Press, 2012.
- Martín Alcoff, Linda. "On Judging Epistemic Credibility: Is Social Identity Relevant?" In *Women of Color and Philosophy: A Critical Reader*, edited by Naomi Zack. Oxford: Blackwell, 2000.
- Matsuda, Mari J. *Words That Wound: Critical Race Theory, Assaultive Speech, And The First Amendment*. 1 edition. Boulder, Colo: Westview Press, 1993.
- McGowan, Mary Kate. "Oppressive Speech." *Australasian Journal of Philosophy* 87, no. 3 (September 1, 2009): 389–407. <https://doi.org/10.1080/00048400802370334>.
- . "Responding to Harmful Speech : The More Speech Response, Counter Speech, and the Complexity of Language Use." In *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public*, edited by Casey Rebecca Johnson, 2018. <https://doi.org/10.4324/9781315181189-12>.
- McMillan, John. *The Methods of Bioethics: An Essay in Meta-Bioethics*. Oxford University Press, 2018.
- Meyer, Ilan H. "Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence." *Psychological Bulletin* 129, no. 5 (September 2003): 674–97. <https://doi.org/10.1037/0033-2909.129.5.674>.
- Mill, John Stuart. *Essays on Politics and Society*. Collected Works of John Stuart Mill, v. 18-19. Toronto ; Buffalo: University of Toronto Press, 1977.
- Nielsen, by Laura Beth. "Power in Public: Reactions, Responses, and Resistance to Offensive Public Speech." In *Speech and Harm: Controversies Over Free Speech*, by Ishani Maitra and Mary Kate McGowan. Oxford University Press, n.d.
- Parekh, Bhikhu. "Is There a Case for Banning Hate Speech?" In *The Content and Context of Hate Speech: Rethinking Regulation and Responses*, edited by Michael Herz and Peter Molnar, 37–56. Cambridge: Cambridge University Press, 2012. <https://doi.org/10.1017/CBO9781139042871.006>.
- Petty, Richard E., and John T. Cacioppo. *Attitudes And Persuasion: Classic And Contemporary Approaches*. Boulder, Colo: Routledge, 1996.
- Post, Robert. "Racist Speech, Democracy, and the First Amendment." *William & Mary Law Review* 32, no. 2 (February 1, 1991): 267.
- Rawls, John. *Political Liberalism*. Columbia University Press, 1996.
- Scanlon, Thomas. "A Theory of Freedom of Expression." *Philosophy and Public Affairs* 1, no. 2 (1972): 204–226.
- Schauer, Frederick. "Uncoupling Free Speech." *Columbia Law Review* 92, no. 6 (1992): 1321–57. <https://doi.org/10.2307/1122997>.
- Song, Sarah. "The Liberal Tightrope: Brettschneider on Free Speech." *BROOKLYN LAW REVIEW* 79 (2014): 12.
- American Civil Liberties Union. "Speech on Campus." Accessed June 8, 2020. <https://www.aclu.org/other/speech-campus>.
- "U.S. Senate Judiciary Committee Hearing on Judge Samuel Alito's Nomination to the Supreme Court." Accessed June 13, 2020. <http://www.washingtonpost.com/wp-dyn/content/article/2006/01/10/AR2006011000781.html>.
- Volokh, Eugene. "Freedom of Speech and Workplace Harassment Comment." *UCLA Law Review* 39, no. 6 (1992 1991): 1791–1872.
- Waldron, Jeremy. *The Harm in Hate Speech*. Cambridge, Massachusetts London, England: Harvard University Press, 2014.

Official Church of Scientology: What is Scientology? “What Is the Significance of the IRS Ruling Regarding Churches of Scientology?” Accessed August 26, 2020.  
<https://www.scientology.org/faq/church-funding/significance-of-irs-ruling.html>.  
Whitney v. California, 274 U.S. 357 (1927).  
Yong, Caleb. “Does Freedom of Speech Include Hate Speech?” *Res Publica* 17, no. 4 (July 13, 2011): 385. <https://doi.org/10.1007/s11158-011-9158-y>.