

# Journal of Cognition and Neuroethics

## The Mental & Physical Health Argument Against Hate Speech

**John Park**

California State University, Sacramento

### **Biography**

Content

### **Publication Details**

*Journal of Cognition and Neuroethics* (ISSN: 2166-5087). February, 2023. Volume 9, Issue 1.

### **Citation**

Park, John. 2023. "The Mental & Physical Health Argument Against Hate Speech." *Journal of Cognition and Neuroethics* 9 (1): 13–34.

# The Mental & Physical Health Argument Against Hate Speech

John Park

## **Abstract**

Overall, there's a rich literature on free speech and hate speech. However, there's been comparatively less discussion on hate speech that brings in empirical psychological and medical evidence on the possible health harms hate speech can have for minorities. I introduce and piece together a set of pre-existing scientific data that's new to the philosophical literature to help sufficiently establish an argument that governments should ban hate speech. Given the adverse effects hate speech can have on one's mental and physical health, hate speech causes harm at many times. Since it causes such harm and leads to overridingly negative consequences, the government ought to regulate such speech.

## **Keywords**

Empirical Moral Psychology, Hate Speech, Free Speech, Political Philosophy

## **1. Introduction**

The issue of free speech occupies an important place in political philosophy. The importance of this issue is even more pressing given the somewhat recent rise of the alt right movement in the U.S. and increase of hate speech on the internet. The question is not whether speech should be limited by the government. Philosophers generally agree that speech should be limited for instances like inciting violence, libel & slander, lying that there's a bomb on a plane, false advertising, blackmail, advertising dangerous adult products to children, child pornography, uttering "fighting" words, lying by yelling 'fire' in a crowded theater, etc. Free speech is not a basic infallible value. The real question is how much should speech be limited. A particular controversial issue that is of singular interest here is whether hate speech should be limited by the government. We can define the term 'hate speech' generally as public communication expressing abusive prejudice and/or violence towards a person or group based on race, religion, gender, ethnicity, disability, age, or sexual orientation.

There are divergences in political regulations on hate speech even within Western liberal democracies. For example, the United States is on the upper end of the spectrum of putting more weight in favor of free speech rather than limiting hate speech.

Meanwhile, places like Germany, the U.K., Poland, France, Hungary, and Austria have hate speech laws in significant part due to the history of Nazis, the Holocaust, and World War II on European soil. Others like Canada, Australia, and Mexico also have hate speech laws. Indeed, regarding hate speech, the U.S. is more of the exception in allowing for it rather than the rule.

Casting a long shadow over the hate speech debate is Mill's (1859/1978) classical work on liberty and the harm principle. The harm principle in light of hate speech says that we are allowed to generally speak freely but harming others provides a pro tanto reason for regulating speech. Mill appears to maintain that harming others is not a sufficient reason for regulating speech. In order to have a sufficient reason, further justification is needed for limiting speech. Mill, who is a consequentialist, says that one must show that the consequences of the harm outweighs the consequences of the benefits of free speech in the given instance. It's this generally accepted or orthodox approach in ethics of one means for when speech and liberty should be limited that's grounded in the harm principle that will be the main mechanism used for our inquiry as to whether hate speech should be regulated by the government.

Moreover, Mill states that the harm principle can be applied prospectively to help prevent future harm. The risk of harm allows for the invocation of the harm principle even if the harm may not come about. For example, he writes in *On Liberty* that we may limit the liberty of individuals from being intoxicated at work since it elevates the risk of workplace accidents and physical harm to fellow employees. In modern times, the government may limit the ability to drive while intoxicated given the risk of bodily injury to others even though no accident may occur in a particular instance. Such limitations are justified even though no harm actually does ensue on a case.

What is a 'harm' is a controversial issue debated by philosophers. While this issue would constitute its own separate paper, we may continue our inquiry for our purposes by relying on a current highly influential understanding of 'harm' given by Joel Feinberg (1984). 'Harm' is not meant to refer to any kind of damage to any object whatsoever, but it is about a wrongful setback to the interests of a person.<sup>1</sup> 'Interests' are one's stake

---

1. I focus on Feinberg's account of harm as his is an influential view in contemporary philosophy. Mill differs in that he defines 'harm' as a setback of one's important interests *in which one has rights*. 'Rights' for Mill ultimately are grounded in utility and promote good consequences. Mill believes that we have a right to be free from unwarranted bodily injury from others. For example, he states that if the buying of poison had no other purpose than to murder others, then it should be banned. He says that drinking on the job should be banned because one might hurt others at work. Insofar as I show that hate speech can cause physical injury to others, I also demonstrate that hate speech is a harm on Mill's account.

on an issue, where one's life may go better or worse in the long run depending on what happens. Feinberg presents examples of interests like career development, keeping one's property, maintaining the well-being of one's family, having social justice for one's community, and having one's country be economically secure. An example of what is not a harm is when one's favorite sports team loses a game. This isn't a setback of one's interests as it is merely a temporary discomfort or pain. If something negative will go away and leave an individual as they were previously, whole and undamaged, then one's interests were not set back. Yet, perhaps there might be fanatics who do become mentally ill in the long term because their team didn't win the championship. Even so, the opposing team beating one's team isn't in-itself morally wrongful. It's just fair athletic competition. Therefore, it isn't a harm. Here, we see the importance of 'wrongful' in an analysis of 'harm.'

Feinberg writes:

[A]n affront or an insult normally causes a momentary sting; we wince, suffer a pang or two, then get on with our work, unharmed and whole. But if the experience is severe, prolonged, or constantly repeated, the mental suffering it causes may become obsessive and incapacitating, and therefore harmful. (1984, 45–46)

Feinberg does not classify hate speech as a harm but as a mere offense as he believes it has just a temporary sting on minorities without any lasting mental or physical damage (1985). Minorities are still left whole in the long term after being insulted with racial slurs, such as when black people are called the N-word. However, notice in Feinberg's quote that if certain wrongful speech is found to have long term effects on mental health, then such speech is a harm. The harm principle comes into play. It will be one goal among others in this paper to demonstrate that wrongful hate speech does at many times cause damage to one's mental and physical health. Therefore, once I demonstrate this, we can state unequivocally that hate speech is a harm.

Mill justifies liberty in speech for many reasons. Some of them are that we should have a marketplace of ideas because a silenced opinion can be true, a silenced opinion importantly may contain a partial truth, and speech that challenges what is true allows us to test the truth so that the truth is not held in mere prejudice that's inherited rather than adopted. However, Mill allows for restrictions on freedom of speech in part with the harm principle. Today, Mill's harm principle-based route is perhaps the predominant means in philosophy for the government potentially limiting liberty and certain speech, although there may be alternative means for doing so as well.

There are numerous contentions against hate speech (Langton 1990, Fish 1994, Waldron 2012, McGowan 2012, Maitra 2012, Brown 2015). For example, Laura Beth Nielsen (2012) draws on empirical studies to show that the proper response to hate speech is not to allow it and advocate for more speech that counters hate speech. For, data demonstrates that minorities subject to hate speech overwhelmingly desire to keep silent after being subjected to hate speech due to numerous factors such as fear from physical harm. Caroline West (2012) contends that hate speech should be banned because empirical data shows that it serves to generally undermine and silence speech rather than enhance it.

Others like David Boonin (2011) argue against regulating hate speech. For example, he says that at many times hate speech can be viewed as fighting words. Since we already have regulations on fighting words, we don't need new regulations on hate speech. Moreover, as we shall continue to see shortly, it's the received view in philosophy that hate speech isn't a health harm and doesn't cause long term damage to health. The harm principle can't be invoked for this reason. Therefore, it's questionable as to whether there needs to be government involvement on it.

While there may be many good arguments above against hate speech, our purpose here is not to assess or expand upon any of such previous contentions against hate speech. Instead, it is to solely focus on a different avenue concerning the effects hate speech has on mental and physical health. Through this route, I contend that hate speech ought to be banned.

## **2. Criticisms of Other Attempts To Use Health Harms**

There have been a few scholars who have discussed the mental effects and health harms hate speech may have in order to argue against hate speech. However, these only have been very small in number, and they only have provided speculative or insufficient cases. I now offer my own criticisms of them. Richard Delgado (1993) speculates that hate speech along with institutionalized discrimination can cause adverse immediate mental effects, such as fear, stress, and anger. However, they might actually have long term effects. They may impact health in that they may lead to mental illness, such as depression, or physical illness, such as high blood pressure. Nevertheless, Delgado's essay lacks rigorous empirical data to establish the health effects hate speech might have, especially given that empirical psychological work on hate speech and discrimination in the early 90's was still a budding field.

Melina Bell (2021) makes a recent attempt to show that hate speech can cause psychological harm in order to contend that hate speech should be banned. However, it is insufficient. She cites studies where women expressed a greater state of self-objectification after viewing comedy skits that objectified women as sex objects (Ford et al. 2015). She then states that self-objectification can lead to mental illness. However, one concern is that these were not longitudinal studies. While the jokes may have caused self-objectification in the short term, it may have been short-lived because the women may have at least subconsciously thought later that it's really only a joke. Professional comedic settings where statements usually are made in jest may provide a different circumstance in the minds of women in the long run. Many women may not stay worked up about it over time, and many may develop mental illnesses over gender discrimination from non-comedic contexts rather than from comedic ones. This is a reasonable possibility, so replicated longitudinal experiments are needed to show that self-objectification from comedy skits contributes to long term health issues rather than only hate speech that's given in non-comedic contexts. Without such studies, Bell's contention is invalid.

Moreover, it also could be the case that relevant speech in comedy skits may cause self-objectification at least initially since others are laughing and agreeing. Meanwhile, hate speech in non-comedy settings might cause no illnesses. One can't necessarily infer from these studies in comedic contexts that hate speech in non-comedy settings causes mental illnesses in many women. Given the context of comedic settings, studies on the effects of discriminatory speech in non-comedy settings are needed in order to sufficiently and decisively show that hate speech in non-comedy settings causes harm.

Bell also relies on studies suggesting that discrimination leads to health problems like heart disease, cancer, and strokes for African Americans (Lewis & Van Dyke 2018). The studies importantly were able to weed out factors like socioeconomic status, health behaviors, and access to care as factors for causing such problems. This makes it more likely that discrimination is the cause of such maladies. However, the studies don't eliminate other important possible confounding variables that may be responsible for the illnesses rather than discrimination. For example, birthplace and education level were not eliminated as possible causal influences for the negative health outcomes. More data is needed to eliminate such possible confounding variables in order to have a valid argument that discrimination and hate speech cause such health illnesses. Overall, Bell is unable to sufficiently show that hate speech is a harm and causes health illnesses. To note, in my own argument below, I will present studies that eliminate the above confounding variables.

In a *Stanford Encyclopedia of Philosophy* entry on freedom of speech, David van Mill notes that regarding hate speech and a Neo-Nazi march in Skokie, Illinois:

[W]e might want to claim that [the minorities] were psychologically harmed by the march. This is much more difficult to demonstrate than harm to a person's legal rights. It seems, therefore, that Mill's [harm principle] does not allow for state intervention in this case (2017).

There's a general sense in philosophy that the empirical evidence for hate speech causing mental illnesses is not established as of yet, or at the very least, it's controversial.

Works in philosophy advocating for an end to hate speech by taking the avenue of medical health have yet to put forth a systematic and sufficient case that hate speech has significant negative effects on mental or physical health. Thus, I will attempt to be the first in the philosophical literature to rely on modern empirical findings in psychology and medicine to sufficiently establish such negative outcomes and then argue for legislative measures against hate speech. One of my contributions is to introduce to the philosophical literature on hate speech the sufficient empirical evidence of how hate speech can cause negative health outcomes. I then argue that hate speech generally should be banned by the government in part given the effects it can have on human health. This is a general moral rather than legal claim I'm making here that the government ought to regulate hate speech, and I leave the later question of what specific policy measures should be implemented by the government to ban hate speech for a later time. I only make a general moral claim here that hate speech ought to be banned.

### **3. Empirical Psychology Evidence on Hate Speech & Health**

The first step of positing my contention against hate speech is to establish that hate speech is a statistically significant factor for causing negative mental and physical health consequences for many minorities. This is a causal empirical claim that hate speech causes health harms. It's not a constitution claim that hate speech itself is a harm, such as when a president utters a discriminatory mandate that itself enacts policy that is harmful. In this instance, the speech constitutes harm since it enacts a harmful policy. While I don't deny the possibility that hate speech also can constitute harm, the focus of my argument is only on a specific causal claim that hate speech can cause harm; namely, to one's health. This causal rather than constitution claim will be what I need to establish my thesis.

Social scientists and researchers in the medical field have amassed hundreds of studies demonstrating that hate speech and other kinds of discrimination can have adverse effects on mental and physical health. Discrimination is a broader category of which hate speech is a token. To note, I rely on studies that specifically involve hate speech along with experiments that involve many other kinds of discrimination. The results are similar for both. However, as hate speech is a type of discrimination, replicated negative health outcomes found for diverse kinds of discrimination provide inductive support that they will apply to hate speech as well. Inductive logic is the logic of probabilities, where a good inductive argument with true premises will support the conclusion with a sufficient degree of likelihood rather than by 100% certainty. If many different kinds of discrimination cause bad health outcomes, then this increases the likelihood that hate speech is causally efficacious too in leading to such outcomes. This is like how discovering that many diverse species of primates practice some level of cooperation provides good reason to believe that a disparate newly discovered primate species also practices cooperation. As many different mRNA-based vaccines in history have proven to be safe in the long term, this means that the new COVID-19 mRNA vaccines, such as by Pfizer and Moderna, will likely be safe in the long term despite an absence of longitudinal studies on the new vaccines. This basic use of inductive logic for my conclusion is further strengthened based on the fact that studies to be discussed below that directly test for the effects of hate speech also are shown to lead to similar health outcomes. Insofar as replicated evidence on hate speech is also utilized, this staves off the objection that hate speech doesn't cause health maladies because experiments just on discrimination that don't involve hate speech can't be used to make such an inference. Moreover, the use of inductive logic, as described above, utilizing other diverse kinds of discrimination also sufficiently addresses this objection.

The existence of hate speech along with data showing that institutionalized discrimination for practices like job hiring (Quillian et al. 2017), housing (Schneider 2018), and criminal justice (Sentas 2018) are still going strong provides an overall environment that can be deleterious to the health of many minorities. It can lead to chronic stress that causes serious health harms.

First, there is psychological evidence that hate speech and discrimination correlate with stress (Krieger & Sydney 1996, Ren et al. 1999, Finch et al. 2000, Gelber & McNamara 2016). Many of these studies are interviews where hate speech and discrimination seemingly co-occur with stress. However, they as of yet don't establish causation that hate speech causes stress. One can't establish causation from mere personal phenomenological reports and psychological interviews since hidden subconscious mental



events that really drive mental processes may be different from conscious awareness and recollection. Rather, more rigorous experimental methods are required to help establish causation. Regardless, such correlational studies are important and potentially still help to contribute to a causal claim because you can't have causation between two events unless you have correlation.

As an example of a correlational national study, African Americans, Hispanic Americans, and Asian Americans reported higher levels of stress than whites supposedly in significant part due to experiencing racism (Williams 2000). In another study, 96% of participants who experienced racism within a past year reported feeling stressed from the incident (Klonoff and Landrine 1999). Nielson (2012) conducted an ethnographic study interviewing those who experienced hate speech. She found that many reported experiencing stress and fear for their safety from such incidents. For instance, a Filipino woman was interviewed and recalled an episode.

And I was at a gas station, and a guy came out, didn't talk to me directly, but I knew he was talking about me. I was seated in the car, and the driver who was beside me was white, and the guy just said, kind of in the air, "I can see the driver's the only human being around here." Implying I was not a human being.

Q: Uh huh. And did you respond to that in any way?

A: No, I didn't because I was afraid. (2012, 156)

In another case, a young woman was interviewed.

I know just last week, I was in the BART station at Montgomery and there was, um, I think a homeless man who came up to me and said, "I hate women, they're all sluts"... That probably sticks in my mind the most...

Q: Um, what did you say to the guy who, um, informed you that all women are sluts?

A: Um, I just turned around; I didn't say anything. I was pretty scared of him. (2012, 161)

A different study based on observation from a psychologist found an association between psychological trauma with an instance of racial discrimination.

A light-skinned Hispanic male was treated courteously when he made application for an apartment in New York City. However, when he returned with his African-American wife, the renting agent became aloof and informed them that the apartment was rented. In response to the denial of the apartment, the wife immediately became depressed, insomniac, and hypervigilant. She had repeated nightmares. At the time of the alleged discrimination, she noticed that her hair had begun to fall out, that her skin was dry, and she was constipated...there was a mild paranoid trend. All of her symptoms were causally related to the discrimination. (Butts 2002, 338)

In a meta-analysis of 333 published articles which involves a total sample across all studies of 309,687 participants, experiencing racism was associated across a variety of minority groups with mental health issues, such as depression, anxiety, posttraumatic stress disorder (PTSD), and suicidal ideation (Paradies et al. 2015). The results were statistically significant. Furthermore, it was significantly correlated with poorer physical health.<sup>2</sup> They write that “[t]his meta-analysis indicates that racism is significantly related to poorer health...” (Paradies et al. 2015, 24).

What’s interesting about this meta-analysis of studies including others (Lewis & Van Dyke 2018) is that they also provide support for a causal claim that hate speech can cause mental and physical health ailments given that many other possible causal factors for the health ailments were eliminated. Through statistical analysis, the researchers found that factors like age, sex, birthplace, socioeconomic status, health behaviors, and education level didn’t moderate the effects of racism on health. It looks like experienced racism rather than other factors is what’s really causing health problems. Paradies et al. (2015, 28) draw a causal conclusion that “racism has long-term effects on health that remain significant despite attenuation over time.” They also point out that given our knowledge of childhood psychology, exposure to racism like hate speech early in life can make one more vulnerable to its health effects with more severe and persistent health consequences given the biological embedding of early life stress.

Moreover, there’s a well-established and uncontroversial literature in psychology and medicine on Adverse Childhood Experiences (ACE). By also eliminating other possible confounding factors, they decisively show that certain negative traumatic childhood experiences like witnessing violence or facing abuse from 0 to 17 years of age can cause

---

2. This correlation fell under the category of miscellaneous physical health.

early mortality, chronic health problems like heart and liver disease, and mental illnesses in victims (Anda et al. 2002, Dong et al. 2004, Brown et al 2010). Part of what's included in ACE is hate speech and racist experiences (Pachter et al. 2009, Wade 2014, Bernard 2021). The ACE studies in themselves establish a strong causal connection of hate speech causing adverse mental and physical health outcomes in youths.

As another study that helps to establish causation, perceived racism by African American women throughout their lifetime accurately predicts their infants' birthweight beyond the effects of medical and sociodemographic factors (Dominguez et al. 2008). Low birthweight babies commonly have complications in fighting off infections, breathing, staying warm, and feeding to gain weight. They are at elevated risk for cerebral palsy, blindness, and deafness. Given that experienced racism is such an accurate predictor of birthweight and many other potential causal factors for decreased birthweight have been eliminated, this suggests that racist acts like hate speech have a deleterious effect on the health of relevant infants. Moreover, suicide rates for ethnic immigrant groups in the U.S. are significantly predicted by the degree of negativity of hate speech they experience (Mullen & Smyth 2004).

Hate speech also can lead to negative physiological changes (Krieger 1990, Krieger and Sidney 1996, Clark et al. 1999, Guyll et al. 2001, Brondolo et al. 2003, Williams et al. 2003, Harrell et al. 2003, Bennett et al. 2004). Physiological changes include activation of the hypothalamic pituitary adrenal cortical system, which then leads to the release of cortisol, a stress hormone. It also entails alterations in immune and cardiovascular functioning. Cardiovascular functioning as well as respiratory functions and pupil size come under control of the sympathetic branch of the autonomic nervous system during emergencies. These physiological changes make up the sympathetic adrenal medullary axis of the stress response, where laboratory and survey evidence link such responses to subjects who experience discrimination (Brondolo et al. 2003, Harrell et al. 2003).

For example, in psychophysiological investigations, subjects experience racially charged encounters in laboratory settings, where physiological effects are measured (Morris-Prather et al. 1996, Kinzie et al. 1998, Blascovich et al. 2001). The racially charged encounters include situations of experiencing hate speech among other discriminatory acts. These are compared to non-racist encounters as controls. Such experiments permit the drawing of cause and effect conclusions about the relationship between hate speech and physiological changes. From overviews of psychophysiological experiments, Harrell et al. state "that direct encounters with discriminatory events contribute to negative health outcomes (2003, 243)."

A moderated psychophysiological approach uses personality measures as predictors of disparities in physiological responses between subjects to racist events. If relevant personality trait differences reliably lead to divergences in physiological responses, then this provides causal evidence that hate speech causes negative physiological harm in relevant participants. For instance, experiments show that those who are more passive in addressing the discrimination they experience have higher blood pressure than those who are more active (Krieger and Sidney 1996). African American women with passive coping responses were 4.4 times more likely to have hypertension or high blood pressure than African American women with active strategies (Krieger 1990). Moreover, diastolic blood pressure reactivity to a speech stressor was augmented for participants who had previously experienced discrimination as compared to those who didn't (Guyll et al. 2001). Hypertension itself is a serious condition that is a leading cause of heart attacks and strokes.

As an alternate argument, the general stress literature can be utilized. Given that hate speech causes stress for minorities, being subject to such speech along with institutionalized discrimination has serious consequences for one's health. At this point in the psychological literature, it's uncontroversial that hate speech can be a very stressful event, where a majority of those who experience hate speech feel stress (Kessler et al. 1997, Clark et al. 1999). It's uncontroversial that hate speech is a significant cause of stress, and it's present within societies with institutionalized discrimination. This environment leads to chronic stress for many individuals. Once we have this on the table, we then can rely on the general stress literature to show that this likely will cause many to suffer poorer health. The harmful effects chronic stress can have on mental and physical health is well documented and uncontroversial. The National Institute of Mental Health states:

[Chronic stress] can disturb the immune, digestive, cardiovascular, sleep, and reproductive systems...Over time, continued strain on your body from stress may contribute to serious health problems, such as heart disease, high blood pressure, diabetes, and other illnesses, including mental disorders such as depression and anxiety. (2020)

This somewhat indirect route that relies on the general stress literature also can be used to establish and buttress the causal claim of the effects hate speech can have on health, although we have discussed above more direct studies showing that hate speech causes many to suffer some of these serious health maladies.

The above kinds of studies help to show the negative influence hate speech has on health. Given, the many replicated studies that eliminate other possible causal variables that can negatively impact health, ACE studies, psychophysiological investigations, moderated psychophysiological experiments, the stress literature, and correlational findings, I believe it's safe to conclude that hate speech is a statistically significant factor for leading to harmful mental and physical health outcomes.

To be sure, whether one experiences damage to one's health will depend on the individual. For example, persons with different ways to cope with hate speech can have different mental and/or physical outcomes. Factors like genetics, personality, childhood experiences such as experiencing hate speech early on, and support networks such as having family or being part of a religious group, can affect what health outcomes one will have through experiencing institutionalized discrimination and hate speech throughout one's life. However, the data still shows that hate speech and other acts of discrimination are statistically significant causal influences for certain negative health outcomes. It's important to keep in mind that even though damage to health can be contingent upon the experiencer of hate speech, those who commit the discriminatory act still do damage to another's health and are causally responsible. This is somewhat similar to how an assaulter hitting another is causally responsible for the ensuing physical harm even though the assailed wasn't dexterous enough as some others to dodge the attack.

I take it that there's a general consensus on the effects of hate speech in the scientific literature. The Surgeon General reports that "the findings indicate that racism and discrimination are clearly stressful events...Racism and discrimination adversely affect health and mental health" (U.S. Department of Health and Human Services 2001, 38) of minorities and places them at risk for mental disorders. In law, given the scientific data, claims of psychological distress in reaction to hate speech have been viewed as credible injuries (Neu 2008).

#### **4. The Mental & Physical Health Argument Against Hate Speech**

Now that we've established that hate speech is a statistically significant cause of adverse health effects, I will contend that this helps to show that hate speech ought to be banned by the government. Recall that harms are wrongful setbacks to one's interests, and this includes long term damage to one's physical and mental health. We have seen how hate speech is a setback in that it can lead to long term damage of one's mental and physical health in non-comedic settings. Furthermore, uttering hate speech against others

is in-itself wrong to do as it expresses abusive discriminatory insults or violence to certain groups. It is a kind of unjust discrimination disparaging or threatening people based on morally arbitrary factors like the color of one's skin. Somewhat similar to how a child is wrong for name-calling on the playground, hate speech is a wrong. Given the above, we can utilize the generally accepted route that's based on the harm principle. We can say as a first pass that speech that causes harm runs into the harm principle, and this provide a pro tanto reason to ban such speech.

Given my new establishment in the philosophical literature that hate speech can cause health injuries, it could be that many kinds of speech may be stressful to others and may cause health harms, such as possibly telling your children to study hard and do well in school. However, telling your children to study hard isn't in-itself wrong to do. Thus, it isn't a harm. Also, recall that when invoking the harm principle, we need to do a consequentialist weighting of the harms and benefits in order to potentially be sufficiently justified in regulating speech. Let us call (1) the application of the harm principle and (2) this consequentialist analysis that lies in favor of regulating action *the harm principle conditions*. Satisfaction of the harm principle conditions is sufficient for banning certain speech. With this in mind, the greater overall good for society is for people to tell their children to study to produce more productive future citizens, so the second condition also isn't satisfied in the case of telling your children to study. Given the above, instructing your children to study shouldn't be banned by the government.

As I already have established the first component that hate speech invokes the harm principle, let us now assess the second component of the harm principle conditions. We need to examine whether the overall consequences favor the benefits of being able to utter hate speech or whether they give more weight to the harms of minorities suffering from death, mental illness, and physical ailments caused by hate speech.

First, the benefits of hate speech are very low. Discriminatory policies can be presented in a civil manner rather than hurling insults and/or physical threats at others. Mill's benefits of free speech concerning uncovering the truth, a partial truth, and making sure we don't hold beliefs dogmatically can be upheld by discriminatory people as they can present their ideas in a civil tone. Uttering hate speech is supererogatory to being free to get across ideas. Hence, Mill's benefits of free speech cannot be used to support hate speech in a consequentialist calculation. In a court of law for a trial, both sides can get across their ideas and arguments without throwing personal insults and/or physical threats to violence at the other party, where judges *will* regulate such abusive vituperations. Likewise, discriminatory people can present their ideas without uttering abusive or threatening hate speech.

We have seen how the benefits of hate speech are minimal. On the other hand, the harms of many people suffering from death, mental illnesses, and physical ailments due to hate speech are tremendous. We know that chronic stress at many times leads to negative physical health outcomes. Recall that hate speech is a statistically significant cause of a variety of physical health illnesses, like having unhealthy infants whose lives are threatened and hypertension which is a leading cause of heart attacks and strokes. Suffering from mental illnesses like depression, PTSD, and anxiety is painful. Mental illnesses can have negative effects on mobility, on the ability to perform physical tasks or even get up out of bed, on one's job performance that impacts subsistence for one's family, on family and friends relationships, on personal well-being, and on pursuit of one's goals in life. Mental health illnesses also can lead to death such as by way of suicide.

Given the little benefits of being able to utter hate speech and the strong negative effects on health of many of those who experience hate speech, the greater consequences are to ban hate speech. We now have seen how hate speech violates the harm principle and how the greater consequences are to ban hate speech. Both components of the harm principle conditions have been satisfied. Therefore, I conclude that the government ought to regulate hate speech. The main contribution to the literature of my essay is that I'm the first to sufficiently argue for the general banning of hate speech by 1) introducing into the philosophical literature a wealth of modern empirical findings that together sufficiently establish that hate speech is harmful to one's health and 2) contending that such health harms in 1) outweigh the little benefits of being able to utter hate speech.

One might object that if we ban hate speech, this will lead to a slippery slope where we'll ban many other different kinds of speech. Soon we'll be living in an authoritarian type of state that has very limited speech such that we can't even criticize the government, and there won't even be freedom of the press. However, unless one establishes that it's causally likely that we'll slip down this slope to an authoritarian-like state concerning speech, one has committed the slippery slope logical fallacy. The burden of proof is on the objector to establish such a causal connection.

Nevertheless, let me point out that there's a stopgap in that my contention against hate speech falls within the harm principle conditions. The harm principle conditions are the stopgap here. I agree that speech that satisfies Mill's harm principle conditions should be outlawed by the government, and this is really as far as I need for my thesis. Therefore, this still perfectly allows for speech that criticizes the government, and it allows for freedom of the press since the harm principle conditions generally don't apply to such speech. Given this stopgap of the harm principle conditions, there's no further reason to think we'll slip down the slope.

Another possible counter is that there is a less costly alternative to government regulation on hate speech that can eliminate the health harms, so there shouldn't be the relevant ban. We should support victims of hate speech to give more speech that counters the hate speech they experienced. This might act as a stress relief such that they don't experience negative health effects. However, as already mentioned, data shows that minorities are not likely to participate in such counter-speech due to factors like fear and intimidation (Nielson 2012, West 2012). After all, historically, there is a credible link between certain instances of hate speech and acts of physical violence against minorities.

David Boonin (2011) argues that there's no need for legislation on hate speech at all. Hate speech is not immediately threatening towards physical violence and isn't immediately "fighting" words because if an elderly and frail white grandmother in a wheelchair said such words, they wouldn't be physically threatening nor stir a minority to fight. Thus, in this respect, there can't be a blanket ban on hate speech that, for example, restricts the grandmother. Also, in contexts where hate speech is threatening towards physical violence or is a "fighting" word, then we already have regulations on such speech, so no new laws are needed. Ultimately, there should be no laws put forth restricting hate speech. However, I have shown that hate speech causes harm to one's mental and physiological health and leads to an overall disutility. Even such words uttered by a frail grandmother can have this effect. Since the harm principle conditions are satisfied, there should be a general legislative ban on hate speech that even covers cases like the frail grandmother.

As another possible objection, one may say that my argument might then allow for banning non-hate speech that causes damage to one's mental or physical health. What about other kinds of unethical speech whose effects are identical to that of hate speech? For example, maybe there should be rules regulating making fun of people who are shorter than average. My response to this is that this might be true, and I would be perfectly happy with the result if so. My contention against hate speech is based in part on the empirical data of the health effects hate speech can have. The fact that there's institutionalized discrimination also exacerbates the situation. This all creates a culture and environment for chronic stress, which leads to negative health outcomes. Moreover, such harms outweigh the benefits. It's in part an empirical question of socio-psychological research whether such an environment or something like it also holds for those who are shorter than average such that they live in a culture that supports chronic stress for them. If the harm principle conditions are satisfied for speech insulting people who are shorter than average, then this provides good reason to outlaw such speech given that it satisfies the harm principle conditions.



An objector may claim that certain speech in the press that causes stress for a criminal might then have to be outlawed on my account, but it shouldn't. For example, on my view, news sources may not be able to print or post articles and editorials condemning a convicted serial child rapist and making it known that an agent is such a criminal. For, this may cause chronic stress for the rapist which can lead to negative health consequences. However, remember the harm principle conditions and my handling of the possible stress caused by telling one's children to study hard. There can be several independent reasons for justifying freedom of the press in this instance. However, for our purposes, it's sufficient to state that writing news stories on crimes isn't in-and-of-itself wrongful. This is especially so when satisfying the normative proposition that the press shouldn't divulge the names of the rape victims in order to protect them. There should be rape shield laws. Since writing news stories on crimes isn't wrongful, this doesn't qualify as a harm. Moreover, it will be for the greater good for society to mete out deserved punishment for such individuals with public rebukes, to inform the public of dangerous people, and to make it known unequivocally that such criminal behavior is undesirable and should not happen. The benefits of such speech from the press outweigh the potential mental harm to the criminal. Thus, my account allows for news sources to print or post accurate articles of and editorials on the child rapist.<sup>3</sup>

In conclusion, it is highly controversial in philosophy whether hate speech should be regulated by the government or not. My contention against hate speech is the first in the philosophical literature to use sufficient evidence from psychology and medicine that shows that hate speech causes mental and physical illnesses. I have shown how this sufficient medical evidence is largely absent in the philosophical literature, and there's the general current thought in the philosophical literature that it hasn't been proven yet that hate speech causes psychological or physical harm. However, I have demonstrated that hate speech is a health harm, and using such data, I also have contended that the health harms outweigh the benefits of uttering hate speech. Therefore, given the satisfaction of the harm principle conditions, hate speech should be banned by the government, and I have defended this thesis from numerous counters. The U.S. should join most of the developed world and regulate such speech that has inimical effects on mental and physical health.

---

3. There are other objections against hate speech regulations in the literature (Post 2012, Heinze 2016, Strossen 2018). However, I take it that they have been adequately rebutted in the literature, so I don't address them here (Delgado & Stefancic 1996, Brown 2015, Brown & Sinclair 2020).

## References

- Anda, R.F., Whitfield, C.L., Felitti, V.J., Chapman, D., Edwards, V.J., Dube, S.R. and Williamson, D.F. 2002. "Adverse Childhood Experiences, Alcoholic Parents, and Later Risk of Alcoholism and Depression." *Psychiatric Services* 53: 1001–1009.
- Bell, Melina. 2021. "John Stuart Mill's Harm Principle and Free Speech: Expanding the notion of harm." *Utilitas* 33: 162–179.
- Bennett, G.G., Merritt, M.M., Edwards, C.L., and Sollers, J.J. 2004. "Perceived Racism and Affective Responses to Ambiguous Interpersonal Interactions Among African American Men." *American Behavioral Scientist* 47: 63–76.
- Bernard, D.L., Calhoun, C.D., Banks, D.E., Halliday, C.A., Hughes-Halbert, C. and Danielson, C.K. 2021. "Making the "C-ACE" for a Culturally-Informed Adverse Childhood Experiences Framework to Understand the Pervasive Mental Health Impact of Racism on Black Youth." *Journal of Child & Adolescent Trauma* 14: 233–247.
- Blascovich, J., Spencer, S.J., Quinn, D. and Steele, C. 2001. "African Americans and high blood pressure: The role of stereotype threat." *Psychological science* 12: 225–229.
- Boonin, D. 2011. *Should Race Matter? Unusual Answers to the Usual Questions*. New York: Cambridge University Press.
- Brink, David. 2018. "Mill's Moral and Political Philosophy." *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/mill-moral-political/>. Last visited 4/30/21.
- Brondolo, E., Rieppi, R., Kelly, K., Gerin, W. 2003. Perceived Racism and Blood Pressure: A Review of the literature and conceptual and methodological critique. *Annals of Behavioral Medicine* 25: 55–65.
- Brown, Alexander. 2015. *Hate Speech Law*. New York: Routledge.
- Brown, A. and Sinclair, A. 2020. *The Politics of Hate Speech Laws*. New York: Routledge.
- Brown, D.W., Anda, R.F., Felitti, V.J., Edwards, V.J., Malarcher, A.M., Croft, J.B. and Giles, W.H. 2010. "Adverse Childhood Experiences are Associated with the Risk of Lung Cancer: A prospective cohort study." *BMC Public Health* 10(1): 1–12.
- Butts, H. 2002. The Black Mask of Humanity: Racial/ethnic discrimination and post-traumatic stress disorder. *Journal of the American Academy of Psychiatry and the Law* 30: 336–339.

- Clark, R., Anderson, N.B., Clark, V.R., Williams, D.R. 1999. "Racism as a Stressor for African Americans: A biopsychosocial Model." *American Psychologist* 54: 805–816.
- Delgado, Richard. 1993. "Words that Wound: A tort action for racial insults, epithets, and name calling. In *Words that Wound*. Edited by M. Matsuda, C. Lawrence, R. Delgado, and K. Crenshaw. Boulder, CO: Westview Press: 89–110.
- Delgado, R. and Stefancic, J. 1996. "Ten Arguments Against Hate-Speech Regulation: How valid?" *Northern Kentucky Law Review* 23: 475–490.
- Dong, M., Giles, W.H., Felitti, V.J., Dube, S.R., Williams, J.E., Chapman, D.P. and Anda, R.F. 2004. "Insights Into Causal Pathways for Ischemic Heart Disease: Adverse childhood experiences study." *Circulation* 110: 1761–1766.
- Feinberg, Joel. 1984. *The Moral Limits of the Criminal Law*. Oxford: Oxford University Press.
- Feinberg, Joel. 1985. *Harm to Others*. Oxford: Oxford University Press.
- Finch, B.K., Kolody, B., and Vega, W.A. 2000. "Perceived Discrimination and Depression Among Mexican Origin Adults in California." *Journal of Health and Social Behavior* 41: 295–313.
- Fish, Stanley. 1994. *There's No Such Thing as Free Speech...and it's a good thing too*. New York: Oxford University Press.
- Ford, T.E., Woodzicka, J.A., Petit, W.E., Richardson, K., and Lappi, S.K. 2015. "Sexist Humor as a Trigger of State Self-objectification in Women." *Humor* 28: 253–269.
- Gelber, Katharine and McNamara, Luke. 2016. "Evidencing the Harms of Hate Speech." *Social Identities* 22: 324–341.
- Guyll, M., Matthews, K.A., Bromberger, J.T. 2001. "Discrimination and Unfair Treatment: Relationship to cardiovascular reactivity among African American and European American women." *Health Psychology* 20: 315–325.
- Harrell, C., Hall, S., Taliaferro, J. 2003. "Physiological Responses to Racism and Discrimination: An assessment of the evidence." *American Journal of Public Health* 93: 243–248.
- Heinze, Eric. 2016. *Hate Speech and Democratic Citizenship*. Oxford: Oxford University Press.
- Kessler, R.C., Mickelson, K.D., and Zhao, S. 1997. "Patterns and Correlates of Self-help Group Membership in the United States." *Social Policy* 27: 27–46.

- Kinzie, J.D., Denney, D., Riley, C., Boehnlein, J., McFarland, B. and Leung, P. 1998. "A cross-cultural study of reactivation of posttraumatic stress disorder symptoms: American and Cambodian psychophysiological response to viewing traumatic video scenes." *The Journal of nervous and mental disease* 186: 670–676.
- Klonoff, E., and Landrine, H. 1999. "Cross Validation of the Schedule of Racist Events." *The Journal of Black Psychology* 25: 231–254.
- Krieger, N. 1990. "Racial and Gender Discrimination: Risk factors for high blood pressure?" *Social Science Medicine* 12: 1273–1281.
- Krieger, N. and Sydney, S. 1996. "Racial Discrimination and Blood Pressure: The CARDIA study of young black and white adults." *American Journal of Public Health* 86: 1370–1378.
- Langton, Rae. 1990. "Whose Right? Ronald Dworkin, Women, and Pornographers." In *Philosophy and Public Affairs* 19: 311–359.
- Lewis, T.T. and Van Dyke, M.E. 2018. "Discrimination and the Health of African Americans: The potential importance of intersectionalities." *Current Directions in Psychological Science* 27: 176–182.
- Maitra, Ishani. 2012. "Subordinating Speech." In *Speech and Harm: Controversies over free speech*. Edited by I. Maitra and M.K. McGowan. Oxford: Oxford University Press, 94–120.
- McGowan, Mary Kate. 2012. "On 'Whites Only' Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination. In *Speech and Harm: Controversies over free speech*. Edited by I. Maitra and M.K. McGowan. Oxford: Oxford University Press, 121–147.
- Mill, David van. 2017. "Freedom of Speech." *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/freedom-speech/>. Last Visited 2/9/2020.
- Mill, John Stuart. 1859/2001. *On Liberty*. Ontario, Canada: Batoche Books Limited.
- Morris-Prather, C.E., Harrell, J.P., Collins, R., Leonard, K.L., Boss, M. and Lee, J.W. 1996. "Gender differences in mood and cardiovascular responses to socially stressful stimuli." *Ethnicity & Disease* 6: 123–131.
- Mullen, B., and Smyth, J. 2004. "Immigrant Suicide Rates as a Function of Ethnophaulisms: Hate speech predicts death. *Psychosomatic Medicine* 66: 343–348.
- National Institute of Mental Health. 2020. "5 Things You Should Know About Stress." <https://www.nimh.nih.gov/health/publications/stress/index.shtml>. Last visited 1/31/2020.

- Neu, J. 2008. *Sticks and Stones: The philosophy of insults*. New York: Oxford University Press.
- Nielsen, Laura B. 2012. "Power in Public: Reactions, responses, and resistance to offensive public speech. In *Speech and Harm: Controversies over free speech*. Edited by I. Maitra and M.K. McGowan, 148–173. Oxford: Oxford University Press.
- Pachter, L.M. and Coll, C.G. 2009. "Racism and Child Health: A review of the literature and future directions." *Journal of Developmental and Behavioral Pediatrics* 30: 255.
- Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A. 2015. "Racism as a Determinant of Health: A systematic review and meta-analysis." *PLoS One* 10, e0138511.
- Post, Robert. 2012. "Interview with Robert Post." In *The Content and Context of Hate Speech*. Edited by M. Herz and P. Molnar, 11–36. Cambridge: Cambridge University Press.
- Quillian, L., Pager, D., Hexel, O. and Midtbøen, A.H. 2017. "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time." *Proceedings of the National Academy of Sciences* 114: 10870–10875.
- Ren, X.S., Amick, B., and Williams, D.R. 1999. "Racial/ethnic Disparities in Health: The interplay between discrimination and socioeconomic status. *Ethnicity & Disease* 9: 151165.
- Schneider, V. 2018. "Racism Knocking at the Door: The Use of Criminal Background Checks in Rental Housing." *U. Rich. L. Rev.* 53: 923.
- Sentas, Vicki. 2018. "Beyond Media Discourse: Locating Race and Racism in Criminal Justice Systems." In *Media, Crime and Racism*, edited by Monish Bhatia, Scott Poynting, and Waqas Tufail, 359–79. Cham: Springer International Publishing.
- Strossen, Nadine. 2018. *Hate: Why we should resist it with free speech, not censorship*. Oxford: Oxford University Press.
- U.S. Department of Health and Human Services. 2001. *Mental Health: Culture, Race, and Ethnicity – A Supplement to Mental Health: A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services.
- Wade, R., Shea, J.A., Rubin, D. and Wood, J. 2014. "Adverse Childhood Experiences of Low-Income Urban Youth." *Pediatrics* 134: e13–e20.
- Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.

- West, Caroline. 2012. "Words that Silence? Freedom of expression and racist hate speech." In *Speech and Harm: Controversies over free speech*. Edited by I. Maitra and M.K. McGowan, 222–248. Oxford: Oxford University Press.
- Williams, D.R. 2000. "Race, Stress, and Mental Health." In *Minority Health in America*. Edited by C. Hogue, M. Hargraves, and K. Scott-Collins, 209–243. Baltimore: Johns Hopkins University Press.
- Williams, D.R., Neighbors, H.W., and Jackson, J.S. 2003. "Racial/ethnic Discrimination and Health: Findings from community studies." *American Journal of Public Health* 93: 200–208.