BEYOND ONE'S OWN MASTERY:
ON THE NORMATIVE FUNCTION OF HATE SPEECH

BEYOND ONE'S OWN MASTERY: ON THE NORMATIVE FUNCTION OF HATE SPEECH

By BIANCA M. WAKED, HONOURS B.A. (FIRST CLASS)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements

for the Degree Master of Arts

McMaster University MASTER OF ARTS (2018) Hamilton, Ontario (Philosophy)

TITLE: Beyond One's Own Mastery: On the Normative Function of Hate Speech

AUTHOR: Bianca M. Waked, Honours B.A. (McGill University) SUPERVISOR: Dr. Wilfrid Waluchow NUMBER OF PAGES: vii, 112.

ABSTRACT

This thesis calls for a reconfiguration of hate speech as a primarily normative phenomenon. All hate speech strives to weaken the social-moral normative status of its targets and in doing, justifies violence against its target. In light of this normative function, the harm of hate speech is reconsidered. Against traditional defenders of hate speech regulation, I claim that individual and collective harm is a highly likely, but not a necessary consequence of hate speech, while intrinsic harm and reckless risk necessarily follow from hate speech's normative capacity. In light of the normative origin of such harms, a societal response with normative clout is required. However, while individual responses are insufficient to block the normativity of hate speech, I suggest that the legal system is characteristically well-suited to do so.

ACKNOWLEDGEMENTS

For supporting me with (very few) questions, I must thank my family. Mom and Dad, you've taken every one of my wayward decisions with grace and patience. I am so proud to be your daughter. Khrystine and Matt, thank you for consistently opening your house —every visit was filled with delicious food and terrible television, making it always so difficult to leave. And to the cleverest of younger sisters—Catherine, it has been so hard being away from the other half of my soul. Thank you for indulging all the phone calls, for the care packages, and for the many visits. Superb creature, indeed.

For their unwavering friendship, I would like to thank Kara, Josh, Rand, Ryan, Kim, Sarah and Natalie. I am constantly in awe of the depth of your kindness, compassion, and strength. A special thanks to Matt, who heard countless iterations of these ideas and berated me accordingly. And of course, I must thank Marten, who has been my constant companion in all (mildly) mischievous social events. The magic of Hamilton, I suspect, largely stems from the magic of these wonderful people.

For reading drafts of every chapter, I must thank Joel. Your love, your support, and your terrible dancing brought smiles when it seemed impossible. My days are always just a little bit brighter with you (and Marlo) in it.

Violetta, you were so often a voice of reason that I needed and I appreciate every moment of it. Your time, effort, and guidance over these past two years was a gift, so thank you. And thank you for allowing me to spend time with Ella, that angel you call a puppy.

And finally, to Wil.

If only there were words.

Expressing the depth of my love and gratitude is, I think, impossible. I have cherished every moment as your student, and I will continue to wear that title proudly. So thank you for our countless discussions walking to classes, thank you for indulging my love affair with exclusive legal positivism, and thank you for sharing your stories. But most of all, thank you for taking me as your student—I have become a better philosopher and a better person for it.

TABLE OF CONTENTS

INTRODUCTION

In the landmark 1992 *Butler* decision, the Supreme Court of Canada ruled that there was "a sufficiently rational link between the criminal sanction, which demonstrates our community's disapproval of the dissemination of materials which potentially victimize women and which restricts the negative influence which such materials have on changes in attitudes and behaviour, and the objective."[1] In recognizing such a rational link, the Supreme Court recognized the importance of legal intervention in protecting vulnerable groups from the harms of such material. The conclusions of this controversial decision, however, find awkward bearing in the philosophical literature surrounding hate speech and the right to freedom of expression. Attempts to explain the "rational link" recognized by the Supreme Court of Canada fall short, if at all, when it comes to the harm of hate speech.

This thesis will therefore investigate hate speech's normative capacity in order to shed light on the manner in which hate speech can harm which, in turn, will clarify whether hate speech regulation is necessarily at odds with the right to freedom of expression. Beginning with John Stuart Mill's *On Liberty* and moving on to Thomas Scanlon's "A Theory of Freedom of Expression", the first chapter sets aside all mentions of hate speech and exclusively considers arguments defending the importance of the right to freedom of expression, while also noting the argumentative possibilities which allow

[1] *R.v. Butler,* [1992] 1 SCR 452, 55-56.

for restrictions on certain types of expression. The following chapter considers early arguments in favour of hate speech regulation defended by critical race legal theorist Mari Matsuda. “Public Responses to Racist Speech: Considering the Victim’s Story” prioritizes the experience of targets, noting the deep physical and psychological trauma which arises as a consequence of such speech. But Matsuda sustains a problematic premise concerning hate speech as merely vitriolic content—an assumption undermined in the third and final chapter. The third chapter of this thesis therefore rejects this premise and emphasizes the normative function of hate speech: weakening the social-moral normative status of targets in order to justify otherwise morally impermissible behaviour. The success of hate speech’s normative function, I argue, is drastically increased in cases wherein the target is a historically vulnerable group, but intrinsic harm and risk are present regardless of the speech act’s success. Individual harm and collective harm, on the other hand, are highly likely to occur, but not strictly necessarily. This failure to distinguish necessary from likely harms is to blame for much of the confusion behind the harm of hate speech, and clarifying this distinction permits further insight into the question of legal regulation.

Taken in conjunction, I conclude that such harms are sufficient to warrant some kind of societal response, but do not necessitate legal intervention *per se.* Even so, I suggest that legal intervention is characteristically suited to strip hate speech of its normative capacity.

This thesis therefore stands as philosophical support for the perceptive conclusions of *R.v. Butler*, a decision which recognized and defended the normative capacity of hate speech.

CHAPTER I

ON THE FREEDOM OF SPEAKERS

Prior to my analysis of hate speech and its position respective to the freedom of expression, I would like to set the scene, so to speak, and clarify the relevant philosophical terrain surrounding the right to freedom of expression. I will begin with John Stuart Mill's arguments in *On Liberty* as well as Thomas Scanlon's arguments posited in his aptly titled article, "A Theory of Freedom of Expression". In proceeding with the arguments supporting a generalized freedom of expression, I endeavour to both explain the force of such arguments as well as highlight the relevant premises which will be later scrutinized in light of the phenomenon of hate speech.

I.i. John Stuart Mill on the Liberty of Thought and Discussion

It is important to note that John Stuart Mill's 1859 publication, *On Liberty,* concerns the broader political project of identifying when state restrictions on individual liberty can be justified. Having said that, the freedom of expression plays an important role in this larger political domain. As such, Mill offers a careful argument defending the only condition which he claims may legitimately be used by the state to justifying individual liberties. "That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their

number", Mill tells us, "is to prevent harm to others."[2] That is, state intervention in the liberties of its individual members can only be justified on the grounds of protecting other members of the society from harm. The full implications of this principle, however, can only be adequately conceptualized once the conclusion has been situated in Mill's larger argument.

On Liberty opens with the claim that social and civil liberties are Mill's main focus of discussion, and not the metaphysical "liberty of the will" that has been so frequently a topic of philosophical discussion. More specifically, Mill is concerned by the tension which exists between the civil liberties granted to individuals in society, and the authority of the governing state.[3] Given that political regimes have developed, the relationship that the individual maintains to the governing state has necessarily altered. This, in turn, entails that the tension between the liberties of the individual and the authority of the state has altered and subsequently "requires a different and more fundamental treatment."[4] Rather than consider the tension which has previously existed between a sovereign ruler or class of individuals and those over whom they ruled, political theory must now contend with the threat of the "tyranny of the majority".

As Mill explains, establishing democratic regimes assuaged the possibility of a tyrant ruler because the state's power "was but the nation's own power, concentrated, and

---

[2] John Stuart Mill, *On Liberty* (Kitchener, Ontario: Batoche Books Limited, 2001), 13.

[3] *Idem,* 6.

[4] *Ibid.*

in a form convenient for exercise."[5] However, practical application of democratic ideals brought to light considerations which had been previously hidden from view. Mill tells us that "the will of the people", which was supposed to be instantiated in the democratic state and subject to its desires, was little more than "the majority, or those who succeed in making themselves accepted as the majority" and that those individuals "consequently may desire to oppress a part of their number."[6] As a consequence of this emerging form of tyranny, which put the majority in conflict with weaker groups, political theory is now tasked with the additional objective of guarding against this abuse of power as well as offering precautions which would mitigate such harmful consequences.

Yet this new adapted form of tyranny is particular in a way that previous forms of tyranny were not. Mill tells us that "when society is itself the tyrant…its means of tyrannizing are not restricted to the acts which it may do by the hands of its political functionaries."[7] That is, the tyranny of the majority does not oppress others merely through political action. Instead, it oppresses dissenting minorities through what Mill calls a "social tyranny", which he takes to be a more powerful tactic of oppression. Unlike political oppression, social tyranny "leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself."[8] This is to say that social tyranny permeates and restricts one's life in a way that political tyranny rarely

---

[5] *Idem,* 7.

[6] *Ibid.*

[7] *Idem,* 9.

[8] *Ibid,*

does. Insofar as it is an oppression which is instantiated through the people, it can be weaponized by individuals surrounding the dissenting minority and affords little opportunity for escape. If the majority of one's social circle belong to the majority, then the social oppression permeates one's everyday life. Escaping such social oppression therefore entails trying to escape one's everyday life in society, which is a difficult, if not outright impossible, task.

Moreover, it is important to note that Mill does not hold political oppression and social oppression to be necessarily incompatible. Rather, social oppression works in conjunction with political oppression in order to not only restrict one's political actions, but also their social well-being in everyday life. Therefore, protection against the tyranny of the majority requires protection from both the political mandates which may be enacted as well as the social pressure which might be utilized against dissenters.

As such, Mill concludes that protection merely against the political mandates of the state is insufficient to guard against the tyranny of the majority, given that it fails to protect against the conditions of social oppression. Instead, "there needs protection also against the tyranny of the prevailing opinion and feeling; against the tendency of society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them."[9] Therefore, the state must institute measures which can protect individuals against both political as well as social tyranny.

---

[9] *Ibid.*

Yet one only has to look at the variety of rules of conduct and customs scattered across history to recognize that determining this limit is easier expressed in theory than in practice. “No one, indeed, acknowledges to himself that his standard of judgment is his own liking,” Mill explains, “but an opinion on a point of conduct, not supported by reasons…and if the reasons, when given, are a mere appeal to a similar preference felt by other people, it is still only many people’s liking instead of one.”[10] Relying on personal preference is a dangerous line of thought, Mill tells us, insofar as it makes it complicates one’s ability to consistently hold society to accountable to legitimate and illegitimate restriction. As such, he formulate a principle which could be accepted by various individuals, each of whom maintaining different conceptions of the good life. Put in slightly different, anachronistic terms, the principle Mill formulates is meant to be content neutral insofar as it does not rely on a particular conception of the good life in order to be politically useful.[11]

Mill’s eventual conclusion, that it is only to prevent harm to others that society may justifiably restrict another’s liberties, is therefore a manner in which states might mitigate tyranny. It should be noted that Mill claims his principle is “entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control, whether the means used be physical force in the form of legal penalties, or the

[10] *Idem,* 10.

[11] *Idem,* 9.

moral coercion of public opinion."[12] This suggests that the harm principle applies to both political and social tyranny, and as such, is meant to serve as both a moral and political principle.

I.ii. The Argument for State Legitimacy

Thus, we may now turn to Mill's arguments regarding freedom of expression with a better understanding of their political function. Entitled "Of the Liberty of Thought and Discussion", Mill's second chapter offers two arguments in favour of freedom of expression, roughly categorized as the instrumental argument and the intrinsic argument. While the first argument concerns the function of freedom of expression in legitimizing the state's power, the second concerns the intellectual and rhetorical benefits of freedom of expression in and of itself.

Mill's state legitimacy argument begins from the premise that freedom of expression is one of the necessary conditions for a free and democratic society, noting later that its protection must be both "absolute and unqualified."[13] Freedom of expression is a necessary condition for the legitimacy of the state insofar as it allows the populace to express dissenting opinions.[14] However, it should be stressed that it is the mere possibility of free expression which legitimizes the state, not the necessity of acting upon such

---

[12] *Idem,* 13.

[13] *Idem,* 15-16.

[14] *Idem,* 18.

freedom. In order to defend this claim, Mill posits a state which shares unanimous opinions with its populace. Apart from sharing the same opinions, this state also never utilizes its power to coerce its people. Even if it were the case that this state could come to be, Mill nevertheless argues that "the power itself is illegitimate", meaning that the state cannot be recognized to be a legitimate governing body.[15] This is due to the fact that it is the possibility of dissent, rather than dissenting opinion itself, which constitutes the grounds of democratic legitimacy. The fact that the governing body welcomes alternative opinions rather than using its power to suppress such alternatives is sufficient to legitimize said government's power. As such, freedom of expression becomes a grounding condition of the legitimacy of the state.

It is worth noting here that an extension of this claim entails that the content of the dissenting opinion is necessarily irrelevant to the question of whether the expression ought to be protected. This is the origin of the content-neutrality clause, a clause later taken up and enforced by legal and political philosophers as a central obstacle to hate speech regulation. While the relationship between content-neutrality and hate speech will be a focus of later chapters, it is important to recognize that the force of the content-neutrality objection—that no restriction on speech can be a direct consequence of the content of said speech—stems from the fact that content neutrality is a necessary condition for democratic legitimacy.

---

[15] *Ibid.*

I.iii. The Argument for Intellectual Enrichment

Furthermore, should we set aside the question of state legitimacy, Mills claims that there remains nevertheless an argument for protecting an unqualified right to freedom of expression. This argument begins from the premise that to silence a dissenter is to rob others of the benefit of engagement, regardless of whether the content of the expression is true or false. Depending on whether the content is true or false, the argument follows two different prongs in order to arrive at an identical conclusion—namely, that society loses an opportunity for intellectual enrichment.

A. The Case of the True Dissenting Opinion

If it is the case that the content of the dissenting opinion is true, then society has lost an opportunity to improve erroneous views. Mill defends this first conclusion by expanding the concept of silencing. Embedded within the act of silencing dissent is an assumption of infallibility in the view commonly accepted. "To refuse a hearing to an opinion, because they are sure that it is false," Mill tells us, "is to assume that their certainty is the same thing as absolute certainty."[16] Insofar as we know that humans are fallible, it is absurd to assume certainty in domains in which we are prone to error. The recognition that there could be alternative views, taken in tandem with the failure to evaluate such views, therefore leaves little ground to justify one's view as correct.

[16] *Idem,* 19.

However, precautions may be taken in light of our fallibility, and Mill supports exposure to a variety of different ideas as one such precaution. More specifically, it is only once possible objections have been considered and alternative views have been evaluated that one can provide sufficient reason in order to justify his chosen view. As Mill explains,

> "The steady habit of correcting and completing his own opinion by collating it with those of others, so far from causing doubt and hesitation in carrying it into practice, is the only stable foundation for a just reliance on it: for, being cognizant of all that can, at least obviously, be said against him, and having taken up his position against all gainsayers—knowing that he has sought for objections and difficulties, instead of avoiding them, and has shut out no light which can be thrown upon the subject from any quarter—he has a right to think his judgment better than that of any person, or any multitude, who have not gone through a similar process."[17]

Our ability to justify our conclusions therefore depends on the possibility of engaging with alternative views. Should it be the case that all other views are discounted in favour of one's original view, there remains nevertheless the additional explanatory power stemming from an engagement with alternatives views and objections, and it is this explanatory power which justifies the reliance on one's judgment.

Finally, it should be noted that this argument is heavily motivated by a contentious premise concerning the possibility of human progress. In addition to the explanatory power which justifies the reliance on one's judgment, Mill also notes that intellectual development and human progress depend on adhering to the logical conclusions of one's reasoning, and this is only possible in an environment where

---

[17] *Idem,* 22.

dissenting opinions might be expressed.[18] Human progress, a notion which concerns Mill throughout *On Liberty*, is only possible when opinions which do not adhere to the orthodoxy are given free reign. If a state restricts the expression of opinions, then the possibility of progress is stifled. However, should this premise be rejected, then the value of dissent is not so obviously tied to the possibility of intellectual improvement.

Baked into the argument against silencing is the relevant, but easily overlooked assumption that the unwillingness to continually analyze one's deeply held premises is necessarily less rigorous. While Mill is right to acknowledge that silencing dissent implicitly entails an assumption concerning the infallibility of one's view, it does not necessarily follow that an unwillingness to continually engage with one's premises produces poorer academic justification. Furthermore, even if we accept the assumption that continuous critical engagement is necessary for proper rigour, disagreement with the claim is not necessary in order to argue that restrictions on certain kinds of speech are justifiable. One could argue that while silencing dissent is the mark of assumed infallibility, there are certain claims of which we *should* be infallibly certain. That is, to subject certain premises to continuous critical analysis might weaken our commitment to such premises and prove more harmful in the long-term. For example, political assumptions concerning equality for all might be one such assumption of which we may claim infallible certainty. The equality clause, the argument might go, is the result of progress in ethical and political thought. To reconsider this conclusion is to run the risk of

[18] *Idem,* 33.

relinquishing such progress given the fraught political history of disenfranchising certain members of the political community.[19] As such, the argument would run, there may be reason to refrain from ascertaining arguments mounted against the equality clause. This type of argument demonstrates the manner in which Mill's argument against the infallibility assumption fails to address the second order question of whether there are claims of which we might be infallibly certain, or more controversially, if there are any claims of which we might *want* to be infallibly certain. While there is more to be said here, later chapters will pursue this line of thought in in greater detail.

B. The Case of the False Dissenting Opinion

Following his argument concerning true dissenting opinions, Mill turns to the possibility that the stifled speech is false. The claim that the speech might be blatantly false is set aside in order to address the stronger possibility that an idea might be mistakenly taken to be true. If it is the case that some idea is obviously false, Mill suggests, then there is little reason to believe that thoughtful individuals will accept and defend such an idea. However, the more epistemically complex case concerns ideas which are not immediately recognized to be false, and as such, Mill's argument considers the consequences which would arise, should such speech be stifled.

[19] I recognize that the equality-equity debate is alive and well in political philosophy, however I am merely using the equality clause as an example of a political premise that ordinarily prompts a pause, should the clause be revoked.

Given that this disjunct of the argument concerns errors of judgment, Mill takes pains to clarify the proper methodology which ought to be utilized in evaluating one's own ideas as well as ideas put forth by others. "There is a class of persons," Mill tells us, "who think it enough if a person assents undoubtingly to what they think is true, though he has no knowledge whatever of the grounds of the opinion."[20] This is a naive approach which, whether true or false, leads to the idea being treated as a kind of belief, independent of argument and evidence. What should therefore be a justified idea subsequently comes to resemble a superstition. Proper understanding of ideas therefore requires more than mere assent to conclusions which seem to be intuitively true—it requires that individuals learn the arguments which motivate their conclusions.[21]

The proper approach to ideas, Mill tells us, is not only relevant for individuals to better reflect on their own positions or beliefs. It is additionally necessary for individuals to reach a conclusion in cases of two or more conflicting sets of reasons which lead to two plausible but different conclusions. As Mill explains, "…on every subject on which difference of opinion is possible, the truth depends on a balance to be struck between two sets of conflicting reasons."[22] Insofar as there will be multiple arguments presented for different conclusions, individuals must be capable of providing a valid argument explaining both why one's preferred conclusion is the proper one and, more importantly,

---

[20] *Idem,* 35.

[21] *Ibid.*

[22] *Ibid.*

why other theories cannot be true. Understanding only the arguments in favour of one's conclusion is insufficient, and "if he is equally unable to refute the reasons on the opposite side; if he does not so much know what they are, he has no ground for preferring either option."[23] Properly understanding the grounds for one's argument therefore includes (i) understanding the grounds which motivate one's preferred arguments, (ii) understanding the alternative arguments in their strongest forms, which justify alternate conclusions, and finally, (iii) understanding the relationship between the different sets of arguments.[24] It is only with these three necessary components can an individual properly understand their conclusion. And should one of these necessary conditions be lacking, Mill claims that the only rational position to take is a suspension of judgment.

This proper understanding is of especial importance when considering commonly accepted knowledge or truisms taken for granted. Commonly accepted knowledge cannot be fully understood until it has been evaluated in light of opposing arguments. "The fatal tendency of mankind to leave off thinking about a thing when it is no longer in doubt," we are told, "is the cause of half [of mankind's] errors."[25] This is not to say that there will be no truisms which remain uncontested, but merely that the process of evaluating ideas is valuable even in light of unanimity with respect to the truth of some truism.

---

[23] *Ibid.*

[24] *Idem,* 35-36.

[25] *Idem,* 41.

Cases in which the populace will be unanimous with respect to the truth, however, are few and few between. Rather, Mill claims that cases in which the different ideas share some truth between them are much more common. Different ideas will share different parts of the truth and omit others. As such, the truth remains fractured among what are often opposing opinions, each of which presenting themselves as the whole truth.[26] The consequence of such a state of affairs is that ideas taken up as true are understood to embody the whole truth, and this comes at the cost of any other truth contained in different or opposing ideas. Contrary to this exclusive conception of the truthiness of ideas, Mill tells us that "every opinion which embodies somewhat of the portion of truth which the common opinion omits, ought to be considered precious, with whatever amount of error and confusion that truth may be blended."[27] This composition of fragmented truth is only possible once we have, as previously mentioned, understood the grounds of the arguments and additionally, come to recognize the truth is unlikely to be found in a single idea.

This conception of the truth, as a quality distributed among different ideas, is the grounding for Mill's defence of dissenting opinions *tout court*. As he notes,

> "When there are persons to be found who form an exception to the apparent unanimity of the world on any subject, even if the world is in the right, it is always probable that dissentients have something worth hearing to say for themselves, and that truth would lose something by their silence."[28]

---

[26] *Idem,* 44.

[27] *Ibid.*

[28] *Idem,* 46.

The possibility of a false dissenting opinion has therefore been renegotiated as the possibility of a dissenting opinion with fragments of truth, and the mere possibility that dissenting opinions might contain some truth is sufficient to justify the necessity of expressing such opinions.

I.iv. On the Authority of Society over Individuals

Mill's argument in *On Liberty* for freedom of expression must be situated in the larger context of the argument concerning society's authority. That is, Mill is concerned with the extent to which society is justified in restricting civil liberties. While Mill does not align his argument with the social contract tradition, he nevertheless supports the notion that society functions on a *quid pro quo* basis—that is, he takes the relationship between individuals and society to be one of reciprocity. "Every one who receives the protection of society owes a return for the benefit, and the fact of living in society renders it indispensable that each should be bound to observe a certain line of conduct towards the rest."[29] This conduct is comprised of two conditions, each of which is necessary to properly fulfill the terms of the relationship. These are as follows:

(i) Prevention of injury to others
(ii) Fair distribution of the burdens of cohabitation

Referred by later scholars as "Mill's Harm Principle", the first condition requires that individuals not harm others by "not injuring the interests of one another; or rather

[29] *Idem,* 69.

certain interests, which, either by express legal provision or by tacit understanding, ought to be considered as rights."[30] Infringements on the interests of others, Mill tells us, is sufficient to justify society wielding its authority to restrict the liberties of culpable individual. Additionally, the second condition requires that each individual bear an equitable share of the burdens associated with cohabitation. Bearing one's fair share of the "labours and sacrifices incurred for defending the society or its members from injury and molestation" is a condition which highlights the necessity of equality among the members of society.[31] Should it be the case that harsher burdens are born by certain groups, then governing bodies are justified in using its authority to alleviate the burden.

However, this inference is complicated by a distinction drawn between acts which *injure* others and acts which merely *inconvenience* others. States cannot justifiably restrict the liberties of an inconsiderate individual who otherwise adheres to the boundaries of basic civil liberties. "The offender," Mill tells us, "may then be justly punished by opinion, though not by law."[32] The inconvenient act or opinions ought to be addressed using non-legal instruments, such as public opinion and education, in the attempt to cultivate less unruly behaviour. Such instruments, Mill tells us, are central to our flourishing in that they enable us "to distinguish the better from the worse" and they offer "encouragement to choose the former and avoid the latter.[33] Yet it is only the

---

[30] *Ibid.*

[31] *Ibid.*

[32] *Idem,* 69.

[33] *Idem,* 70.

definite risk or certainty of damage to others which can justify the intervention of the law, and such conduct must otherwise be tolerated as the price of individual liberty.

Yet the question of equitable distribution brings to the fore what the proper course of action should be in cases to the contrary. Or, put differently, Mill assumes that such inconveniences will be, by and large, even among individuals in saying that such inconveniences are ones which society "can afford to bear."[34] Yet this claim does not exclude the possibility of justified action should there be an unjust distribution of inconveniences. Seeing as Mill does not explicitly consider this question, he offers little guidance as to what exactly would constitute a justified intervention and the degree to which the unfair burdens would have to be born before it breached the threshold and constituted a harm. But it is important to noted that this line of questioning does not come into conflict with Mill's argument, and that the state may, at least in theory, maintain its legitimacy even in correcting for an unjust distribution of inconveniences.

It should additionally be noted that Mill acknowledges groups for which his arguments do not apply. Certain groups may have their liberties curtailed without invoking the justifying conditions. As he explains, "[i]t is, perhaps, hardly necessary to say that this doctrine is meant to apply only to human beings in the maturity of their faculties. We are not speaking of children, or of young persons below the age which the

[34] *Idem,* 76.

law may fix as that of manhood or womanhood."[35] In other words, society may justifiably restrict the liberties of individuals who have not yet reached maturity.

Mill similarly exempts "those backward states of society in which the race itself may be considered as in its nonage," given that he deems such societies to be under developed.[36] "The early difficulties in the way of spontaneous progress are so great," Mill explains, "that there is seldom any choice of means for overcoming them; and a ruler full of the spirit of improvement is warranted in the use of any expedients that will attain an end, perhaps otherwise unattainable."[37] The necessity for society to progress therefore justifies tyranny or other forms of rule which would ordinarily be illegitimate. While it should be noted that a number of these societies are likely non-white, there is no necessity tying such under-developed societies to particular races. As such, Mill's framework is conditional upon certain conditions obtaining, the most important of which being full "maturity of faculties".

And yet, this conjunction leaves Mill's argument in a precarious logical position. Seeing as the argument refuses the possibility of certain individuals having full rational capacity, their experience is omitted from the discussion of free speech and justified restrictions on liberty. Apart from being a generally problematic moral position, this omission is argumentatively worrisome in light of discussion surrounding the harm of

---

[35] *Idem,*13-14.

[36] *Idem,* 14.

[37] *Ibid.*

hate speech, insofar as racial minorities thought to belong to "under-developed" societies are among its most popular targets. Trivially, the force of a universal framework depends on its applicability to all members of society. Insofar as Mill uses the criterion of rational maturity to omit any group, he calls into question the universal applicability of his framework. As such, the framework must at least be brought into conversation with excluded groups in order to pass muster. While I do not take the exclusion to discredit the Millian framework altogether, it minimally introduces a further philosophical step concerning any relevant features which might complicate this model for groups excluded by the maturity condition. Therefore, any consistent defence of Mill's theory must either introduce this philosophical step, or reject its exclusionary premises.

II.i. Scanlon's Defence of Freedom of Expression

The arguments of *On Liberty* undoubtedly provide the framework underlying the conversation surrounding freedom of expression. Yet they are largely a product of their age, as I suggested in the earlier discussion. And yet, political theorists are hesitant to relinquish the powerful insights of Mill's theory. The task for contemporary theorists committed to the Millian framework is therefore to adapt the arguments to the current political terrain. Thomas Scanlon's defence of the doctrine of free speech, in his (aptly) titled article, "A Theory of Freedom of Expression", is one well-known example of such. This theory diverges from Mill's aforementioned racial difficulties by applying his framework across racial lines. Scanlon subsequently relies on the foundations of the

Millian framework while exploring the nature of harm and the category of "expression" to bridge together contemporary jurisprudential practices and theories of free expression.[38]

Scanlon's article aims to answer the charge of irrationality that often follows strong readings of the doctrine of freedom of expression. "On any very strong version of the doctrine," Scanlon explains, "there will be cases where protected acts are held to be immune from restriction despite the fact that they have as consequences harms which would normally be sufficient to justify the imposition of legal sanctions."[39] Insofar as it is the state's task to protect its citizens from harm, intervention in cases of harmful speech seems necessary, yet the doctrine of freedom of expression holds that it is (generally) more important to protect the freedom of individuals to express themselves than to guard against the particular harms of certain forms of speech. It is the tension which exists between these two claims which results in the charge of irrationality that Scanlon seeks to address in his article.

Addressing this charge of irrationality, Scanlon tells us, will require a philosophical justification composed of at least two distinct conditions. Firstly, the theory must be able to provide a proper account of the class of protected speech, that is, the theory must be able to provide us with the means of identifying what forms of speech are

---

[38] While Scanlon exclusively considers the American jurisprudential tradition, his arguments are not exclusively restricted to this body of work. However, it should be noted that he often relies on intuitions which seem to be particularly pertinent to the American tradition.

[39] Thomas Scanlon, "A Theory of Freedom of Expression," *Philosophy and Public Affairs* 1, no. 2 (Winter 1972): 204.

afforded legal protection. Secondly, the theory defending freedom of expression must explain "the nature and grounds of [freedom of expression's] privilege", that is, a proper defence must clarify the justification behind such legal protection of freedom of expression.[40] In providing these two components, any defence of the doctrine of freedom of expression will be capable of answering questions which arise as a consequence of the doctrine. Examples of such include determining to what extent the doctrine of free expression is an artificial consequent of current political institutions and to what degree is it a natural doctrine which exists outside the creation of such institutions.

II.ii. On Concepts and Clarifications

Scanlon begins with a general definition of the class "acts of expression", in which he includes "any act that is intended by its agent to communicate to one or more persons some proposition or attitude."[41] He then considers the possibility that the doctrine of freedom of expression concerns those acts which address a large or public audience and assume that the communicative proposition is of interest to a general audience.[42]

However, Scanlon qualifies this claim by pointing to the fact that this definition paints too broad a stroke and captures more acts than would be preferred. More specifically, it captures what Scanlon calls "violent and arbitrarily destructive" acts and,

---

[40] *Ibid.*

[41] *Idem,* 206.

[42] *Ibid.*

as he explains, "it seems unlikely that anyone would maintain that as a class they were immune from legal restrictions."[43] Too broad a definition captures acts which we would not want protected from legal penalties merely in virtue of the fact that such acts communicate a proposition. It is therefore insufficient to invoke the protection of the doctrine of free expression that an act merely communicate some proposition or attitude, and as such, the law is distinguishing between protected and unprotected acts in some capacity.

One possible grounds for this distinction, Scanlon considers, is the "speech/act" distinction. This distinction rests on a more technical use of the term "speech" in which acts not ordinarily taken to be forms of "speech" are captured by such a term.[44] Miming and written/non-verbal communication are the examples raised by Scanlon that the "speech/act" distinction might have trouble defending—namely, how can typically non-verbal expressions be protected by the law under the "speech" category when there is a category better suited to capture the nature of the act (namely, "acts"), but is not afforded a similar blanket legal immunity? Scanlon rightly identifies the "speech/act" distinction as problematic insofar as it stems from the assumption that all protected acts of expression must share some property through which they are distinguished from other

[43] *Idem,* 207.

[44] *Ibid.*

acts for which we would not want legal protection.[45] "It could be, and I think is, the case," Scanlon explains,

> "[T]hat the theoretical bases of the doctrine of freedom of expression are multiple and diverse, and while the next effect of these elements taken together is extended to some acts a certain privileged status, there is no theoretically interesting (and certainly no simple and intuitive) definition of the class of acts which enjoys this privilege."[46]

Rather than attempt to identify a condition or property which is shared by all acts protected by the doctrine of free expression, Scanlon takes the class of protected acts as a diverse whole, therefore adopting an exclusive disjunctive framework when considering the grounds on which said acts are protected.

This is one of Scanlon's more insightful claims, as far as its implications are concerned. In accepting the possibility of diverse justifications and thereby tailoring a disjunctive framework motivating the doctrine of free expression, Scanlon rejects the possibility of reducing the doctrine of free expression to a single condition which only captures relevant cases. In doing so, he acknowledges that theorists may have to contend themselves with a plurality of justifications, none of which can be axiomatically reduced. Yet in forfeiting the shared condition, Scanlon distances himself from *a priori* determinations regarding the acts afforded legal protection. While accepting multiple justifications for the doctrine of free expression is not equivalent with recognizing the

---

[45] *Idem,* 208.

[46] *Ibid.*

role that context plays in affording acts legal protection, it can be interpreted as a step forward in such a direction.

As such, Scanlon's response to the charge of irrationality will not rely on the identification of a pure class of protected acts. He instead turns to classic violations of freedom of expression in order to identify the commonalities among them. "What distinguishes these violations from innocent regulation of expression," Scanlon tells us, "is not the character of the acts they interfere with but rather what they hope to achieve."[47] Scanlon suggests that intuitions concerning legitimate and illegitimate restrictions on expressions stem from the justification offered for such restrictions. That is to say, the goal motivating restrictions on speech determines the legitimacy of the restriction. Restrictions premised on preventing the distribution of particular ideas seem less legitimate, Scanlon suggests, than restrictions concerned with features of the expression.[48] As such, we might look towards a content/form distinction as the guiding principle of the doctrine of freedom of expression.

And yet, this distinction is not without its problems. The difficulty with determining what exactly constitutes "the view communicated" by an expression might pose more problems than the distinction is worth. Additionally, the content/form distinction delegitimizes defamation laws. Insofar as defamation laws are premised on the notion that "it would be a bad thing if the view communicated by certain acts of

---

[47] *Idem,* 209.

[48] *Ibid.*

expression were to become generally believed", they seem irreconcilable with the claim that content-based restrictions are illegitimate.[49] As Scanlon notes, defamation laws are a prime example of justifiable, content-based restrictions on expression which ought not be excluded by theories. Given that any strong reading of the content/form distinction entirely forfeits defamation laws, it is incapable (in and of itself) of capturing relevant nuances of the doctrine of freedom of expression.

II.iii. On the Consequences of Expressions

The cost of a strong content/form distinction leads Scanlon to consider the manner in which certain acts induce harm, with a focus on "cases where these harms clearly can be counted as reasons for restricting the acts that give rise to them".[50] Scanlon does not think such cases are sufficient to justify restrictions, but merely that they are present regardless of whether the restriction is justified. The list, composed of six separate circumstances in which harm can be induced through expression, is as follows[51]:

(i) Expressions which directly cause harmful physical consequences
(ii) Expressions in which the production (form) of the act necessarily involves the view communicated (content)
(iii) Slander
(iv) A man falsely shouting fire in a crowded theatre
(v) Expressions which indirectly contribute to a harmful act committed by another
(vi) Expressions which radically decreases public safety or significantly increase the ability of individuals to harm one another

---

[49] *Ibid.*

[50] *Idem,* 210.

[51] *Idem,* 210-212.

However, (vi) is offered in two different forms. In the first instance, the expression under consideration is a recipe for homemade nerve gas, while the second case features political propaganda which comparably decreases public safety. Scanlon claims that "in these cases the matter seems to me to be entirely different, and the harmful consequences seems clearly not to be a justification for restricting the acts of expression."[52] The nerve gas, he intuits, could be justifiably restricted from distribution while the political propaganda could not, according to such terms.

Thus, Scanlon builds off his content/form distinction by introducing a further distinction between action-motivation and action-facilitation.[53] As he explains,

> "A person who acts on reasons he has acquired from another's act of expression acts on what *he* has come to believe and has judged to be a sufficient basis for action. The contribution to the genesis of his action made by the act of expression is, so to speak, superseded by the agent's own judgment."[54]

The relevant factor distinguishing the political propaganda from the nerve gas recipe seems to be the individual's judgment. The indirect contribution on the part of another pales in comparison to the individual digesting such reasons and making a judgment. In the latter case, the political propaganda has convinced the individual, and in doing so, has become the individual's motivation for action. However, Scanlon suggests

---

[52] *Idem,* 212.

[53] *Idem,* 212.

[54] *Ibid.*

that the same cannot be said of the individual who provides others with a nerve gas recipe —there is a contribution which exists beyond the presentation of persuasive reasons.

While this distinction touches on questions of complicity and accessory, Scanlon is quick to shy away from any positive claim concerning the nature of complicity. "I am interested only in maintaining the negative thesis", Scanlon tells us, "that whatever these crimes involve, it has to be something more than merely the communication of persuasive reasons for action (or perhaps some special circumstances, such as diminished capacity of the person persuaded)."[55] As such, his claim concerns the insufficiency of persuasion to justify restrictions on expression, and nothing further.

One concern which arises as a consequence of Scanlon's distinction between action-motivation and action-facilitation centres on Scanlon's conception of special circumstances. The question of what constitutes special circumstances such that the restrictions would be more readily accepted is poorly substantiated. Apart from a passing reference to diminished rational capacities as a possible case of special circumstances, we are afforded little direction as to what sort of circumstances can justify restrictions on expression. While Scanlon will consider war and other states of emergency as exceptional circumstances in which the doctrine of freedom of expression might possibly be suspended, there seems to nevertheless be a difference. Temporary states of emergency in which the doctrine is temporarily suspended do not alter the ordinary application of the doctrine of freedom of expression, yet the special circumstances which Scanlon refers to

[55] *Idem,* 213.

here do exactly that. They seem to be a set of conditions which alter the manner in which the doctrine is generally applied.

Scanlon's primary conclusion, however, emphasizes that communication of persuasive reasons is insufficient to justify restrictions on freedom of expression. This negative claim should sound familiar, as Scanlon takes it to be a natural extension of Mill's argument in Chapter II of *On Liberty.* Recall that Chapter II, "Of the Liberty of Thought and Discussion", concerned the legitimacy of the sovereign authority, which was only possible if its constituents were afforded the opportunity to express dissenting and divergent opinions. This was justified on the grounds that silencing the opinion would result in a loss, whether the opinion was true or false.[56] Extending Mill's argument concerning the communication of false belief, Scanlon outlines his Millian principle as the following:

> "There are certain harms which, although they would not occur but for certain acts of expression, nonetheless cannot be taken as part of a justification for legal restrictions these acts. These harms are: (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression; (b) harmful consequences of acts performed as a result of those acts of expression, *where the connection between the acts of expression and the subsequently harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing.*" (emphasis mine)[57]

Scanlon claims that this principle both tracks our intuitions concerning legal responsibility as well as captures the relevant difference between political propaganda

---

[56] John Stuart Mill, *On Liberty* (Kitchener, Ontario: Batoche Books Limited, 2001), 18.

[57] Thomas Scanlon, "A Theory of Freedom of Expression," Philosophy and Public Affairs 1, no. 2 (Winter 1972): 213.

and the nerve gas recipe examples. While not, in and of itself, sufficient to serve as the justification for the doctrine of freedom of expression, it is offered as the fundamental principle underlying the doctrine. This is due to the fact that it explains why certain consequences of expressions are insufficient to justify restriction on said expression and thus, begins to answer the charge of irrationality.[58] Scanlon additionally highlights the fact that the Millian principle does not invoke special rights or appeal to a particular value of certain types of expression, be it artistic, scientific, or political. These two benefits therefore lead Scanlon to introduce the Millian principle as the foundational notion underlying the doctrine of freedom of expression.

If we accept the different intuitions concerning the political propaganda and the nerve gas recipe, then Scanlon's principle successfully tracks the different underlying intuitions. However, it is worth noting that Scanlon's argument is restricted to a particular conception of causal harm—that is, the principle concerns the harm which is a direct causal result of the communication of some expression. In other words, the harm can be tangibly retraced to the communication of some expression. While this conception of causal harm is most familiar, it is far from the only model through which harm might ensue. As later chapters will explore, different kinds of of harm-incitement have gained traction over the past twenty years, in reference to the phenomenon of hate speech, and not all of them rely on direct causation. While the introduction of different kinds of harm-incitement need not undermine Scanlon's model, given his disjunctive framework, a

[58] *Idem,* 214.

proper account of the doctrine freedom of expression must, at the very least, grapple with these emerging accounts of harm-incitement.

II.iv. The Argument from Autonomy

Yet the story concerning the doctrine of freedom of expression is far from over. Observance of the Millian principle is not merely a supererogatory duty to be fulfilled by governments at their own discretion. Like Mill, Scanlon takes it to be a consequence of the argument concerning the legitimacy and authority of governments. As such, the relationship between legitimate governments and individual's self- perception as autonomous depends on governments adhering to the Millian principle.

Scanlon begins his sub-argument with the claim that a legitimate government necessarily must allow for individuals to regard themselves as equal, autonomous agents. The primary condition demands that individuals understand themselves to be sovereign in choosing beliefs as well as evaluating different sets of competing reasons for action.[59] In other words, "an autonomous person cannot accept without independent consideration the judgment of others as to what he should believe or what he should do."[60] This is not to say that individuals are expected to meet the highest standards of rationality, nor is it to say that they may never rely on the judgment of others. Rather, reliance on the judgment of others must be supplemented with independent reasons for thinking that another's

[59] *Idem,* 215.

[60] *Idem,* 216.

judgment would be more likely to arrive at the best answer. Therefore, citizens are afforded the opportunity to arrive at conclusions independently, whether such conclusions are independently arrived at or the result of a second-order decision to defer to an external party.

As Scanlon highlights, this is a minimal condition for autonomy which requires nothing more than independent evaluation and the opportunity to make one's own proper judgments. The weakness of this condition entails that Scanlon's condition is consistent with coercion, particularly on the part of the state. He explains this consistency by explaining that the evaluation of conflicting or forceful considerations remains up to the individual and "[a] coercer merely changes the considerations which militate for or against certain course of action."[61] While certain considerations may be imbued with urgency as a consequence of coercion, this force has little impact on the autonomy of the coerced individual.

The compatibility of coercion and autonomy leads Scanlon to suggest that the state may have a special right to command action of its people.[62] In other words, the fact that the state commands a certain act (X) of individuals provides a strong reason for completing the act. However, the individual remains autonomous in this situation seeing as the recognition that a certain act is required does not entail the performance of said act. It merely offers an additional reason to be considered by the individual.

---

[61] *Ibid.*

[62] *Ibid.*

Yet, the fact that a state's special right to command is not necessarily in contradiction with individual autonomy does not mean that the special right remains unqualified. The special right to command could take a variety of forms, and certain forms contradict Scanlon's condition for autonomy. As he notes, individuals who recognize the state's special right to command "could not regard themselves as being under an 'obligation' to believe the decrees of the state to be correct, nor could they concede to the state the right to have its decrees obeyed without deliberation."[63] Recognizing the necessity of such qualifications is therefore to recognize the necessity of the Millian principle for the legitimacy of the state's authority. Or, put differently, the Millian principle is the requirement for all individuals to continually recognize themselves as autonomous while recognizing the legitimate authority of the state to command action.

Thus, we find ourselves at the heart of "the irrationality of the doctrine of freedom of expression."[64] It is the tension which exists between the simultaneous requirement of the Millian principle, which legitimizes state authority, and the obligation of the state to evaluate threats and invoke legal action when necessary, in order to protect its citizenry. The Millian principle therefore restricts a state's acting capacity according to the following two conditions:[65]

---

[63] *Idem,* 217.

[64] *Ibid.*

[65] *Idem,* 217-218.

(i) The state cannot protect against the harm of coming to have false beliefs.
(ii) The state cannot outlaw the advocacy of conduct which it has outlawed.

The first condition of the Millian principle limits the state by ensuring that individuals maintain their right to independent judgment concerning their beliefs. Insofar as the autonomy of citizens in conjunction with a state determining what beliefs are permissible leads to a contradiction, a legitimate authority cannot support legal intervention according to such terms.[66]

At the risk of suggesting that Scanlon's claim is stronger than it intends to be, it is important to note that Scanlon recognizes that an individual might grant the state the authority to act for him in evaluating beliefs and still reserve the possibility of determining the truth of the matter according to the remaining arguments. However, he claims that this judgment is empty at best. "While he would not be under obligation to accept the state's judgment as correct," Scanlon tells us, "he would have conceded to the state the right to deprive him of grounds for making an independent judgment."[67] As such, even in a case where individuals reserve the right to judge the truth while allowing the state to guide their actions, they have nevertheless ceded their autonomy to the state. The case therefore fails to undermine Scanlon's original argument concerning the tension between autonomy and the state's screening of beliefs.

Moreover, the second condition offers a similar argument regarding the tension between individual autonomy and restrictions on expression. While the first condition

---

[66] *Ibid.*

[67] *Idem,* 218.

sought to protect the right of individuals to make judgments, the second condition works at a second-order level and seeks to protect the right of individuals to access arguments which justify their judgments. To concede the right to restrict the advocacy of previously outlawed conduct to the state is to concede the right to evaluate arguments in favour of the outlawed conduct, which is to concede the right to properly evaluate the conduct altogether. "[This] is a concession that autonomous citizens could not make," we are told, "since it gives the state the right to deprive citizens of the grounds for arriving at an independent judgment as to whether the law should be obeyed."[68] Thus, appropriate evaluation of laws requires that arguments be available to citizens, thereby rendering impermissible a state's decision to restrict arguments.

These arguments should sound familiar, given that Mill defended similar claims concerning the necessity of arguments in order to increase chances of discovering truth.[69] However, Scanlon differentiates his arguments by pointing to the fact that his arguments relies neither on empirical facts concerning the probability of discovering the truth, nor the fact that it would be "an outstandingly poor strategy for bringing about a situation in which true opinions prevail", given the inclinations of human nature.[70] It is instead grounded in limitations on the authority of states over citizens, emphasizing a different, but related angle. The question which has hitherto concerned Scanlon is the following:

---

[68] *Ibid.*

[69] *Ibid.*

[70] *Idem,* 219.

> "Could an autonomous individual regard the state as having, not as part of a special voluntary agreement with him but as part of its normal powers qua state, the power to put such an arrangement into effect without his consent whenever *it* (i.e., the legislative authorities) judged that to be advisable?" (emphasis not mine)[71]

Scanlon's argument for this distinction rests on an example of a person who, upon self-reflection, recognizing that he would be better off relying of the judgment of his friends. Entering into this agreement is neither irrational nor inconsistent with his autonomy, Scanlon tells us, but this conclusion cannot be extended to include states. In other words, there remains a relevant difference between the case of an individual assessing that his friend's judgment would lead to better outcomes and the case of a state including, *among its normal powers,* the capacity to introduce such an agreement into effect without prior consent. It is the latter which proves to be inconsistent with autonomy, and it is therefore the latter which would render a state's authority illegitimate.

However, he grants that there remains a case to be made in allowing states to restrict expression given circumstances in which acting rationally is impossible. This possibility returns him to the fourth types of harm listed, which briefly examined the case of a man falsely shouting fire in a crowded theatre.[72] "Part of what makes the restriction acceptable is the idea that the persons in the theatre who react to the shout are under

---

[71] *Ibid.*

[72] *Idem,* 211.

conditions that diminish their capacity for rational deliberation."[73] The blatant attempt to incite chaos or harm is therefore sufficient to justify restrictions on such expressions.

Yet Scanlon suggests that this example is trivial and does not properly correspond to what is at stake when considering state restrictions on expression. Insofar as the supposed "harm" to be prevented is not subject to any controversy and the diminished rational capacity is brief and evenly applies all subjects to the conditions, it is a restriction which "would receive unanimous consent if that were asked."[74] Furthermore, individuals who are prevented from hearing a false shout are prevented from making a judgement only in the weakest sense. As such, the example lacks the complications which makes legal restrictions of expression so complex.

While he is quick to write off this example as trivial, there are at least three consequences worth nothing which follow from Scanlon's discussion. Firstly, this examples serves as a notable exception to the distinction drawn earlier between motivation for action and facilitation of action. Recall that Scanlon's distinction between the distribution of political propaganda and a nerve gas recipe was justified according to the fact that the former merely motivated action while the latter facilitated it. The man who cried "Fire!" does not neatly track the distinction and suggests that there are examples which cannot be neatly classified as either. If there is at least one example which straddles category lines, it seems likely that there would be others. This heuristic in

---

[73] *Idem,* 220.

[74] *Ibid.*

demarcating justifiable and non-justifiable restrictions on speech is therefore weaker than initially anticipated.

Secondly, seeing as the man falsely shouting fire is deemed a case in which restrictions are acceptable, establishing an analogy between this case and another would allow for the possibility of justifiable restrictions. Rather than an indication of the triviality of the example, however, I take this to be evidence that even the strongest free speech proponents accept some restrictions on expression as justified. It is worth noting that an analogy may be drawn between the case of the man falsely shouting fire and speech communicating pernicious lies such that panic ensues among the general population. Should a strong enough analogy be established, such speech could be justifiably regulated. Such an analogy need not be accepted, but an argument would nevertheless be required to establish the relevant differences between the two cases.

Finally, Scanlon's argument supporting the doctrine of freedom of expression is not indifferent to the truth or falsity of a claim, that is, the content. This fact also reflects the capacity of his argument to account for defamation laws. Scanlon's defence of the Millian principle might rest primarily on limitations of the state's authority, but it nevertheless includes an evaluation of the worth of expressions. When said evaluation is conducted according to standards of truth, we might more readily accept the law's conclusions. However, truth and falsity are not the only standards which might pertain to the restriction of expressions, and they are the least controversial set of standards which might be utilized.

II.v. In Defence of a Mixed Theory

The concerns surrounding the example of the man who falsely shouted "Fire" leaves us with the fact that there remain other cases which fall beyond the periphery of the Millian principle. Furthermore, capturing these cases with only the Millian principle requires such gross distortions. In light of these peripheral examples, Scanlon identifies three conditions which, taken in conjunction with the Millian principle, captures his intuitions concerning violations of freedom of speech. The three conditions are as follows:

(i) Balancing of values
(ii) Equal access to means of expression
(iii) Expression and special rights

The first condition concerns the balancing of the value of expressions in comparison to other social values. Scanlon presents examples of banning posters and handouts for the sake of cleanliness or environmental protection, which might strike us as strange, a reaction that is "a reflection of our belief that free expression is a good which ranks above the maintenance of absolute peace and quiet [and] clean streets."[75] This intuition is taken to suggest that the doctrine of freedom of expression rests upon a balance of goods—that is, evaluations of the value of the expression must be weighed against competing values. Additionally, different types of expressions will weigh more

[75] *Idem,* 222.

heavily than others in the class of "expressions", and therefore cannot be subject to a single evaluation writ large.[76]

However, this claim provokes the question of how different expressions ought to be weighed and whose evaluations should be prioritized. Within minimal restrictions, we are told this evaluation ought to track the popular will of the people. In other words, public opinion should be the determinant factor in evaluating the worth of expressions *in light of competing considerations*. If not for this qualification, Scanlon's claim would make him vulnerable to criticisms concerning the tyranny of the majority. Popular support, the criticism would go, is easily levied by the majority of the population to quiet dissenting voices, yet this possibility is restricted by incorporating the demands of distributive justice.

Scanlon's following condition holds that states must recognize the manner in which "access to means of expression for whatever purposes one may have in mind is a good which can be fairly or unfairly distributed among the members of a society."[77] Notice that this condition stringently shies away from questions concerning the manner in which expressions ought to be received, or unequal distribution of the authority behind such expressions. Thus, we come to Scanlon's second condition to be taken in conjunction with the Millian principle—the requirement that society's members be afforded equal access to the means of expression.

---

[76] *Idem,* 223.

[77] *Ibid.*

Though nearly complete, the picture Scanlon paints requires one more condition. Insofar as access to expression is the necessary condition through which individuals may participate in the political process, states must consider the weightiness of expression in light of such special rights. Scanlon explains that "[a]t the very least the recognition of such rights will require governments to insure that means of expression are readily available through which individuals and small groups can make their views on political issues known, and to insure that the principal means of expression in the society do not fall under the control of any particular segment of the community."[78] Ensuring that regulations on expression recognize the special status of expression subsequently ensures that political legitimacy is sustained. Regulations, apart from respecting the Millian principle and the first two conditions, must therefore attend to the special relationship between expression and special rights, of which political participation is a central example, as a condition of political legitimacy.

These three conditions, taken in tandem with the Millian principle, constitute Scanlon's mixed theory, acknowledged to be "somewhat cumbersome" but also "mutually irreducible and essential" to any inquiry into freedom of expression.

II.vi. Diminished Rationality Reconsidered

Scanlon's paper closes with reconsideration of the case for diminished rationality. There might seem to be a stronger case, Scanlon posits, if we consider times of war and

[78] *Ibid.*

other states of emergencies. The argument for justified restrictions on freedom of speech during emergencies begins from the premise that speech can cause substantial chaos or harm. Given that overcoming states of emergencies as quickly and easily as possible requires minimal chaos, then it is the case that states could be justified in temporarily restricting speech which promotes chaos or harm during such periods.

This argument can, to some degree, be captured by the Millian Principle. Recall the sixth category of harm outlined by Scanlon in Section II. "The Millian Principle," he tells us, "allows one, even in normal times, to consider whether the publication of certain information might present serious hazards to public safety by giving people the capacity to inflict certain harms."[79] Given Scanlon's emphasis on the hazard to public safety, the capacity to cause chaos and harm can be appropriated by the aforementioned category.

While Scanlon is quick to set aside the question of harm in favour of a more elaborate discussion of the authority of states, there is some residual tension in the claim he puts forth concerning harm. More specifically, I am not certain that his sixth form of harm properly captures the example of "publications which might present serious hazards to public safety".[80] In his initial discussion of the sixth form of harm, Scanlon contrasted providing instructions on how to make a homemade nerve gas with the publication of controversial political papers and argued that the former could justifiably be subject to restrictions on expression while the latter would not. However, his later

---

[79] *Idem,* 224.

[80] *Ibid.*

acknowledgement that "risks…worth taking in time of peace in order to allow full discussion of, say, certain scientific questions, might be intolerable in wartime" suggests some degree of ambiguity. Specifically, there exists a gap between the initial example of the publication of information and the latter example of scientific discussions and debate. It was the insufficiency of published opinions to move beyond motivation and facilitate action that distinguished the published political opinions from the nerve gas recipe. Yet, in recognizing the possibility that scientific discussion and debate might prove too costly in states of emergencies, Scanlon is recognizing the extent to which context affects the causal capacities of language. While admittedly not a focal point of this section, this recognition nevertheless highlights that even Scanlon's account of freedom of expression must acknowledge the role of context in altering the causal capacities of language. Scanlon's *laissez-faire* approach to this example therefore disguises the importance of its implications for the causal efficacy of language.

And yet, this example leads Scanlon to distinguish political debates as unique from other forms of expression which the state might justifiably restrict. Insofar as free expression of political ideas is necessary to ensure the continued legitimacy of the state, the state cannot stifle political debates. To do so would be to stifle the very condition of its legitimate authority. While it is possible that it might "be right for certain people, who normally exercised the kind of authority held to be legitimate by democratic political theory, to take measures which this authority does not justify", Scanlon hesitates to

widely incorporate such possibilities into his defence.[81] This hesitation concretizes into a distinction Scanlon makes between at least two different kinds of authority: authority legitimated by democratic political theory, and coercive authority which has not be legitimated by democratic political theory. There exists "a number of different justifications for the exercise of coercive authority", Scanlon tells us, but this authority nevertheless "differs both in justification and extent from that which, if democratic political theory is correct, a legitimate democratic government enjoys."[82] Note that Scanlon does not invoke the language of illegitimacy in levying his claim and as such, recognizes that such an authority might nevertheless be legitimate. Yet the fact is that he remains firm in his acknowledgement that legitimacy might be the cost we pay to renege on the terms of the Millian principle.

II.vii. Concluding Considerations

Scanlon's argument concerning the legitimacy of state authority is powerful, however it remains one of many possibilities. Consider his autonomy condition. While it might be favourable given its weakness of force, it is far from the only candidate. Other conditions are available which do not interfere with the conditions of the Millian principle. Take Joseph Raz's account, for example. While Scanlon does not consider other theories of authority in his discussion, Raz's normal justification thesis provides an

---

[81] *Idem,* 225.

[82] *Idem, 225-226.*

alternative in which states may restrict some expressions without delegitimizing their authority. As such, regulations on expression do not necessarily come at the cost of state legitimacy.[83]

Furthermore, we should be wary of a catch-all term such as "false beliefs" applied broadly by Scanlon. Both Mill and Scanlon suggest that conclusions concerning the truth of a claim should not be left for the state to determine in their usage of such generalized terms. While this stance might be appropriate in certain cases, it will not necessarily hold in all cases—that is, it is not necessarily true that it will always be inappropriate for the state to decide questions of truth. Consider two distinct, but related questions here:

(i) Is it the case that there might be certain beliefs or some ideas which the state *should* take a definitive stance on?
(ii) Will the state be better able to fulfill its coordinative and political function if it were to offer a definitive stance toward some idea?

While Scanlon touches on the first question with his discussion of harm in the second section of his paper, the question is far from exhausted. That is, might there be additional reasons for the state to take a stance on certain ideas, apart from causing harm? Likewise, how does the state's ability to perform its political and coordinative functions apply to the question of state's authority? Barring a brief mention early on concerning the paradoxical appearance of the doctrine of freedom of expression, Scanlon ignores the relevance of this possibility. He instead focuses on the question of state legitimacy which, while important, is not exclusively so. Furthermore, the question of legitimacy can be

[83] For further details, see Joseph Raz, *Between Authority and Interpretation: On the Theory of Law and Practical Reason* (New York, Oxford University Press Inc., 2009).

reimagined according to the function of the state. Namely, how does a state's incapacity (or unwillingness) to address harm targeting particular social groups delegitimate the state? I take the legitimacy of a state to be constituted by a conjunction of both its ability to perform its function of protecting its populace as well as its ability to sustain individual autonomy through its governing practices. Scanlon takes the former to come at the cost of the latter, but I reject the necessity of this price. Or rather, I have not yet been convinced that other argumentative avenues have been sufficiently considered.

Finally, Scanlon and Mill both suffer from a naive conceptions of expressions. That is, they take the communication of ideas through expression to be the most pertinent factor for their discussion. However, we might ask whether certain expressions are performing additional functions in conjunction with promoting a set of beliefs, which might be legally relevant? Moreover, is it appropriate to group together such expressions which might be performing multiply-relevant functions along with the category of expressions which lack such additional functions? And finally, should this difference warrant a legal distinction?

These are the questions motivating Mari Matsuda. Marked by an interest in expression's impact on historically vulnerable groups, she offers an alternative position one might take towards expression and the harm which might consequently arise. The harm of certain expressions, she will argue, might be more complex and more vicious than initially credited by Mill, Scanlon, and other traditional free speech defenders.

Temporarily suspending the concerns of speakers and the legitimacy of state restrictions, I will now turn to the vulnerability of those confronted with hate speech.

CHAPTER II

ON THE VULNERABILITY OF TARGETS

Thus, we begin our analysis of hate speech with a clearer understanding of the arguments motivating the right to freedom of expression. This chapter proposes not to undermine the arguments put forth by Mill and Scanlon, but to expand their purview—that is, explore how such arguments can be brought to bear on the phenomenon of hate speech. With this aim in mind, we turn to Mari Matsuda's1993 article, "Public Response to Racist Speech: Considering the Victim's Story". Matsuda defends both the possibility and necessity of legal regulation for hate speech by drawing the harm caused by racist hate speech into conversation with Millian arguments defending the right to freedom of speech.

I.i. Preliminary Clarifications

Prior to elaborating Matsuda's arguments, however, a quick, preemptory word is needed concerning Matsuda's language in the article. Given that her article is situated within the context of the United States of America, her article explicitly concerns itself with the First Amendment and freedom of "speech" as opposed to "expression". As such, I will adhere to the language of her article, but apply her argument to freedom of expression. The differences between the legal categories of "speech" and "expression", while otherwise relevant, will therefore be set aside.

In a similar vein, I will follow Matsuda in using the term "victim". While I recognize the problematic assumptions which feminists and other social theorists have noted entangled in the term, I will follow Matsuda when discussing her arguments. However, I do not apply this term uncritically.

Furthermore, Matsuda relies on the terms "racism" and "racist" to motivate her arguments. While obviously loaded terms with broad and diverse references, the terms strictly refer to "the ideology of racial supremacy and the mechanisms for keeping selected victim groups in subordinated positions."[84] The terms therefore concern evaluations of the moral quality of institutions according to the their adherence to racial hierarchies and subordination, and nothing further.

Finally, it must be noted that arguments for restrictions on expression are *not* arguments against the right to free speech. Or at least, they need not be. Matsuda takes racist hate speech to be a particular case worthy of exemption. While Matsuda's arguments might be more or less successful at their task, they should nevertheless be taken as arguments for exceptions to the First Amendment, not as arguments against the existence of such rules *tout court.*

I.ii. Mari Matsuda on Outsider's Jurisprudence

---

[84] Mari J. Matsuda, "Public Response to Racist Speech: Considering the Victim's Story," in *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* by Kimberly Williams Crenshaw, Richard Delgado, Charles R. Lawrence III, and Mari J. Matsuda (Boulder, Colarado: Westview Press, 1993), 23.

The First Amendment can co-exist with racist hate speech regulations—this is the global conclusion Matsuda defends in her article. As suggested by her title, she defends this conclusion by balancing the United States' prioritization of the First Amendment with the perspective of those targeted by racist hate speech.[85] "…[A]n absolutist first amendment response to hate speech has the effect of perpetuating racism," she notes, "[and] tolerance of hate speech is not tolerance borne by the community at large. Rather, it is a psychic tax imposed on those least able to pay."[86] This task is complicated, Matsuda tells us, by American jurisprudence's "dangerously fickle collective commitment to freedom of discourse."[87] This "fickle commitment" leads Matsuda to clarify the argumentative context surrounding the First Amendment in order to defend both the possibility and the urgency of hate speech regulation.[88]

Matsuda approaches the analysis of hate speech as a proponent of "outsider jurisprudence"—that is, "a methodology grounded in the particulars of social reality and experience."[89] Outsider jurisprudence understands law to be necessarily political as well as a pragmatic measure to instigate social change. It additionally undermines the "hidden roots" of racisms by attacking its effects. This is not to say that only targeting the effects is sufficient to eradicate racism, but merely that attacking such effects non-negligibly

---

[85] *Idem,* 17.

[86] *Ibid.*

[87] *Idem,* 18.

[88] *Ibid.*

[89] *Idem,* 19.

damages "the mechanisms of racial supremacy."[90] These considerations motivate Matsuda's emphasis on victims as those best positioned to understand the effects of hate speech.

Insofar as Matsuda supports outsider jurisprudence, she accepts the claim that attacking the effects of racism damages underlying racist ideologies. However, one can prioritize the experience of victims as it pertains to hate speech without accepting the additional claim that racist structures can be so damaged. In other words, there is no necessary relationship between prioritizing the experience of victims and targeting the effects of racism as a method to damage racist structures. Emphasizing social reality therefore does not necessitate support for hate speech legislation, in and of itself, even if the two are claims are often taken in conjunction with one another.

Similarly, the invoked metaphor of "hidden roots" is significant, as it establishes hate speech as a consequence of the larger problem of racism. This characterization suggests hate speech to be causally inert—that is, the by-product of racism, the proper problem. While common, this particular conception of hate speech shapes the kind of problem hate speech is taken to be. My concerns with such a conception will be elaborated at length in the following chapter, however it bears worth highlighting the conceptual influence such a metaphor exerts on our conception of hate speech.

I.iii. An Outsider's Perspective on the First Amendment

---

[90] *Ibid.*

Classical thought, Matsuda tells us, disregards *ad hominem* analysis to be little more than a logical fallacy.[91] Yet proper understanding of hate speech requires expanding the discussion surrounding hate speech to include the experience of target groups.[92] While non-target groups often diminish the harm of hate speech, the experience of victims suggest that this harm should be taken seriously. "The typical reaction of target-group members to an incident of racist propaganda is alarm and immediate calls for redress. The typical reaction of non-members is to consider the incidents isolated pranks, the product of sick but harmless minds."[93] There are two possibilities entailed by these opposing perspectives:

(i) The harm of hate speech is characterized such that hate speech is not sufficiently harmful to require institutional intervention.
(ii) The harm of hate speech is characterized such that hate speech is sufficiently harmful to require institutional intervention.

The perspective one takes therefore determines whether hate speech is a problem to be addressed by the law, which in turn determines whether an institutional response is necessary. Thus, the perspective towards the harm of hate speech is central to determining the law's response.

As a consequence of such high stakes, Matsuda further elaborates the two focal orientations to hate speech, which are categorized as follows:[94]

---

[91] *Idem,* 20.

[92] Insofar as Matsuda restricts her arguments to racist hate speech, all uses of the term "hate speech" in this chapter refer only to racist hate speech, unless indicated otherwise.

[93] *Ibid.*

[94] *Ibid.*

(i) Deflation
(ii) Redress.

Members of non-target groups often react along the lines of "Deflation", Matsuda explains, by reducing hate speech to isolated incidents or tasteless pranks. Such reactions entail no need for institutional intervention insofar as deflation leads to one of two conclusions concerning the harm of hate speech: either no harm has occurred, or the harm is insufficient to justify using government authority. In the case of the former, there is a lack of recognition of the harm, which negates the possibility of institutional intervention given that no wrong has been committed. In the case of the latter, however, the harm of hate speech is recognized but fails to meet the threshold required to justify the use of state power against individuals. As such, deflationary reactions towards hate speech do not necessitate a lack of recognition of the harm of hate speech. They may instead be motivated by Millian concerns regarding the immense power of the state being motivated against individuals gone awry. In either case, however, the argument for state intervention does not obtain.[95]

In contrast, members of target groups tell a different story. Members of target groups defend the necessity of "Redress," which requires institutional intervention. This conclusion follows from Outsider jurisprudence's claim that less egregious forms of racism (the aforementioned "effects of the hidden roots of racism") degenerate into more sinister forms of harm, often culminating in violence or genocide.[96] Violence and

[95] *Ibid.*

[96] *Idem,* 23-24.

genocide, racial hate messages, disparagement, threats, overt disparate treatment, covert disparate treatment and sanitized racist comments are therefore all tools of structural racism, each of which facilitate, to varying degrees, violence. As Matsuda notes,

> "Violence is a necessary and inevitable part of the structure of racism. It is the final solution, as facists know, barely held at bay while the racist weapons of segregation, disparagement, and hate propaganda do their work. The historical connection of all the tools of racism is a record against which to consider a legal response to racist speech."[97]

A proper institutional response must therefore recognize the logical connection of the disparate treatment, segregation, and other tools of racism in order to understand the role hate speech plays in racial violence.

Matsuda's claim concerning the logical connection between racial hate speech and other forms of racist treatment should recall Scanlon's discussion of the distinction between speech and acts in the previous chapter. While Matsuda's claim is localized to racist acts, she is nevertheless arguing against the enforcement of this distinction in the case of racist hate speech. Or, at the very least, she is arguing for the legal recognition of the particular connection between racist speech and the violent acts which are taken to follow.

Yet arguments in favour of "Redress" include a further qualification—namely, that it would be deflationary to restrict discussions concerning the harm of hate speech to the incident itself. Far from occurring in a vacuum, the call for redress is a recognition of the manner in which racist mechanisms empower racist hate speech. Public hate

---

[97] *Idem,* 24.

propaganda, de facto segregated schools and community centres, disguised disparate treatment are all forms of racism which targets bear in conjunction with such incidents. Drawing on evidence from the social sciences as support, Matsuda tells us that explicit messages of racial inferiority in hate speech often lead to physical and psychological symptoms of distress, inchoate self-esteem, isolation, rejection of one's identity, and constant fear for personal security.[98] Current social scientific and scientific work largely echoes or strengthens the conclusions of Matsuda's now dated empirical evidence, and the study of transgenerational effects of racism through epigenetics, is one such example.[99]

Additional harms associated with racist hate speech include threats and violence directed towards members of non-target groups who maintain close personal ties and relations with target groups members. These threats complicate tensions between racial groups and serve to implicitly enforce social segregation among racial groups. The harm of racist hate messages therefore traverses the targets and impacts racial relations more generally. Recognizing the full force of racist hate speech, the call for redress goes, therefore requires situating such incidents in their larger context.

---

[98] *Idem,* 25.

[99] For further details, see Kwame McKenzie, "Racism and Health: Antiracism Is An Important Health Issue," *British Medical Journal* 326, no.7380 (January 2003): 65-66. For a philosophical discussion of such health issues, see Shannon Sullivan, "Inheriting Racist Disparities in Health: Epigenetics and the Transgenerational Effects of White Racism," *Critical Philosophy of Race* 1, no. 2 (2013): 190-218.

How, then, is the law to mitigate such drastically different conceptions of hate speech and its harm? Matsuda supports the call for redress, a conclusion which is motivated by the acceptance of at least three premises. Firstly, Matsuda accepts the evidence concerning the significant damage of racist hate speech as including the physical and psychological distress of targets, the threats towards non-target members, and antagonized race relations. She also accepts the further claim that such harms meet the legal threshold required to justify mobilizing the state's power against individuals. Finally, she accepts a third premise concerning the negligible value of racist hate speech.[100] As she explains, "If the harm of racist hate messages is significant, and the truth value marginal, the doctrinal space for regulation of such speech becomes a possibility."[101] These three claims, taken together, lead Matsuda to conclude that state intervention in cases of racist hate speech are justified.

It is important to note that Matsuda's third premise concerning the negligible value of hate speech is an indirect response to Mill's arguments. More specifically, it is a response to Mill's claim that suppression of speech entails a loss, regardless of whether the suppressed speech is true. Yet it is worth recalling that Mill considered truth to be distributed across ideas, suggesting that even false ideas might have grains of truth. "[E]very opinion which embodies somewhat of the portion of truth which the common

---

[100] Matsuda explains the truth value to be "marginal" and I take her claim to concern the social or epistemic value of the speech. While she might possibly be relying on the aforementioned Millian conception of truth shared across ideas, it seems unlikely given that she does not explicitly address his arguments concerning shared truths.

[101] *Idem,* 26.

opinion omits," Mill suggested, "ought to be considered precious, with whatever amount of error and confusion that truth may be blended."[102] Given that Mill understood truth to be scattered across ideas, Matsuda's claim concerning the marginal value of racist hate speech fails to properly address Mill's concern. Even if there is little truth to be found in racist hate speech, the Millian line would run, such truth ought not be suppressed.

Despite this oversight, I do not think it must be granted that to suppress racist hate speech is necessarily to suppress some grain of truth. Rather, one could argue that racist claims are necessarily false and as a consequence, the suppression of such speech does not suppress any truth. Hate speech, it should be noted, are universal claims pertaining to a target group. While they may not always take the explicit form of a universal proposition, they necessarily rely on a universal quantifier in order to motivate the hateful, persecutory message. The necessity of a universal quantifier entails an impossibly high standard for this claim to be true—namely, it requires that the proposition hold true for all members of said target group.[103] If it is the case that the proposition does not hold for a single member of the target group, the universal claim is falsified, thereby making the hate speech a false claim. And one would be hard pressed to defend the notion that a necessarily false claim has some grain of truth. The necessary

---

[102] John Stuart Mill, *On Liberty* (Kitchener, Ontario: Batoche Books Limited, 2001), 44.

[103] This, of course, does not begin to consider the difficulties with determining the members of said target group. Should the claim be racist, then all members of said race would have to be included. But determining the characteristics used to identify a particular race is a highly complex and controversial question. Additionally, biracial individuals would pose a serious question, thereby complicating an already complicated endeavour.

falsity of hate speech, as a practical matter, would have therefore been a more effective argument in response to Mill's sub-argument concerning the suppression of true opinions.

But it should be highlighted that this argument is not an effective response to Mill's sub-argument concerning the suppression of false opinions. Recall that his argument concerned the benefits of engaging with false claims in order to better understand one's own arguments and conclusions. Mill could respond that such false claims are beneficial insofar as they enable individuals to interrogate and further their understanding to their moral commitments. An additional argument must be provided, in conjunction with the necessary falsity of hate speech, in order to properly respond to this sub-argument. Such an argument will be provided in the next chapter. Briefly, however, the argument will highlight the normative effects of subjecting equality claims for target groups to continuous interrogation and suggest that such intellectual rigour comes at a dangerous cost for such groups.

I.iv. International and Domestic Perspectives on Racist Hate Speech

Having argued for the logical possibility of legal regulation of hate speech, Matsuda next considers whether such regulations are practically possible in the American context. "The questions presented here," she explains, "are whether the values of the first amendment are in irresolvable conflict with the international movement toward

elimination of racist hate propaganda."[104] In order to answer this question, Matsuda brings the anti-racist hate speech position adopted by the international community to bear on the American prioritization of the First Amendment.

Establishing hate speech regulation as a practical possibility requires that Matsuda first consider what measures have already been taken by other nations as well as by the international community. At the time of writing, Article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination required that nation-states criminalize racist hate speech.[105] While Matsuda acknowledges the controversy which surrounded (and continues to surround) causal questions concerning incitement to hatred, the proposed convention nevertheless passed, thereby outlawing the mere dissemination of racist ideas.[106] Matsuda uses international decisions restricting hate speech in order to draw attention to the fact that "[the international community] recognizes that avoiding the spread of hatred is a legitimate object of the law."[107] It is not unprecedented for nation-states to regulate expressions which contain elements of "discrimination, connection to violence, and messages of inferiority, hatred, or persecution" when there exists evidence to reasonably suggest a causal connection.[108] The fact that such evidence "reasonably

---

[104] Mari J. Matsuda, "Public Response to Racist Speech: Considering the Victim's Story," in *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* by Kimberly Williams Crenshaw, Richard Delgado, Charles R. Lawrence III, and Mari J. Matsuda (Boulder, Colarado: Westview Press, 1993), 34.

[105] *Ibid.*

[106] *Idem,* 27-28.

[107] *Idem,* 29-31.

[108] *Idem,* 31.

suggests" a causal connection is not to say that no determinate relation exists beyond reasonable suggestion, but merely that it would be impossible to make explicit a determinate causal connection.

Yet Matsuda also respects the relevance of her context, given the United States of America's overwhelming commitment to the doctrine of free speech. American jurisprudence remains heavily inspired by the Millian line—the threat of granting governments the capacity to regulate expression among its regular power, the increased possibility of obtaining the best argument or conclusion, the difficulty of outlawing certain ideas in light of the ever developing intellectual terrain, etc.[109] The arguments raised in the previous chapter find their bearing in the US legal context, and if the state feels threatened by certain ideas, Matsuda notes, "…[i]t is not without recourse. It can use education and counter speech to combat those ideas. It can control conduct or action arising from those ideas."[110] This commitment to the First Amendment, coupled with the acknowledgement of accessible counter-measures to hate speech which do not involve legal regulation, has led many First Amendment scholars to the conclusion that the regulation of hate speech is not only unnecessary, but also dangerous.

Be that as it may, however, the priority granted to the First Amendment is neither unqualified nor without exceptions. Matsuda highlights such exempted categories of speech in order to demonstrate that such exceptions are sufficiently frequent such that it

---

[109] *Idem,* 31-32.

[110] *Idem,* 32.

would not be unthinkable to justify restrictions on hate speech.[111] Commerce, industrial relations, public servants or other governmental employees, and false advertisement are all cases in which the terms of the First Amendment are either overridden or deemed non-applicable.[112] In addition, rights to privacy and protection against defamation, speech threatening public order, obscenity, and threats also remain exempted from the First Amendment. Taken in tandem, these examples demonstrate the necessity of a robust and nuanced conception of the right to free speech.[113] The distinction between protected and unprotected speech is therefore inevitable, and this inevitability leads Matsuda to conclude that the First Amendment is not "in irresolvable conflict" with regulations on racist hate speech.

Drawing attention to exempted categories of speech is a crucial argumentative tactic for defenders of hate speech legislation for at least two reasons. Firstly, it concretely demonstrates the fact that the US legal system acknowledges the necessity of speech restrictions in certain areas, even in light of the prioritization of the First Amendment. Secondly, it facilitates analogical arguments between hate speech and other categories of accepted restricted speech. One such example is the category of obscenity. The extraordinary similarities between the category of obscenity and the category of

---

[111] *Idem,* 34.

[112] While Matsuda notes the varying justifications that motivate such exceptions, they are not germane to her larger point that American jurisprudence already accepts limitations on the right to freedom of speech. Many of these arguments were considered in depth in the previous chapter and as such, will not be restated.

[113] *Idem,* 31-34.

racist hate speech suggest that legal regulation of one and not the other seems inconsistent. While Matsuda herself does not make this claim, analogies between obscenity and hate speech demand additional arguments defending their relevant differences. It might also provoke questions regarding the harm of obscenity and the empirical evidence used to justify such restrictions.[114] Of course, one might argue that legal restrictions on obscene speech are inconsistent with the First Amendment, thereby circumventing the analogical argument altogether. Yet this would be an extreme position to hold, given that this argument requires disavowing a legal category which has, and continues to be, invoked and supported in the US legal system. Though only one such example, analogical arguments between obscenity and hate speech facilitate justifications for hate speech regulations in the United States.

I.v. Defining Hate Speech

Having defended the logical and practical possibility of hate speech regulation in the United States, Matsuda begins her positive account of hate speech by outlining a legal definition of racist hate speech. Insofar as she aims to ease tensions between the First Amendment and hate speech restrictions, she starts from the premise that any definition of actionable racist speech "must be narrow in order to respect First Amendment

[114] For a more detailed discussion of the manner in which the legal category of obscenity has been used to target and subordinate African Americans, see Kimberlè Williams Crenshaw's "Beyond Racism and Misogyny: Black Feminism and 2 Live Crew" in the *Words that Wound* collection.

values."[115] In accordance with this qualification, Matsuda proposes that racist hate speech be considered a *sui generis* category. This is due to her conclusion that any definition of racist hate speech must either stretch the First Amendment conceptual fabric as an exemption, or identify distinctive features of racist hate speech and formulate a unique category.[116] Matsuda is skeptical of the first proposal, insofar as she claims that stretching the existing legal fabric to accommodate racist hate speech would weaken the fabric. It would do so by "creating neutral holes" which would inadvertently remove or weaken protection from other categories of speech. As such, Matsuda rejects this possibility.[117]

On the other hand, Matsuda recognizes that a content-based legal restriction, wherein the content of the speech is used to identify hate speech, calls to light worries against censorship that defenders of the First Amendment staunchly disavow. However, Matsuda claims that racist hate speech is substantively different from other forms of speech, such as Marxist political speech. The relevant difference stems from a shared principle, the truth of which is undoubted—namely, the wrongness of racial supremacy or a racial hierarchy. This shared principle is the condition Matsuda uses to distinguish permissible hate speech regulations from impermissible, otherwise content-based restrictions. As she explains,

> "How can one argue for censorship of racist hate messages without encouraging a revival of McCarthyism? There is an important difference that comes from human

---

[115] *Idem,* 35.

[116] *Idem,* 36.

[117] *Ibid.*

> experience, our only source of collective knowledge. We know, from our collective historical knowledge, that slavery was wrong. We know white minority rule in South Africa is wrong. This knowledge is reflected in the universal acceptance of the wrongness of the doctrine of racial supremacy. There is no nation left on this planet that submits as its national self-expression the view that Hitler was right…We have fought wars and spilled blood to establish the universal acceptance of this principle. The universality of the principle, in a world bereft of agreement on many things, is a mark of collective human progress."[118]

Matsuda therefore argues that a rejection the doctrine of racial supremacy is substantively different from a rejection of Marxism, according to shared anti-racist principles which are the result of a collective historical legacy.

Doctrines of racial supremacy meet this condition while Marxism obviously fails.. Put differently, Marxist speech is not universally condemned and "by its very content it is political speech going to the core of ongoing political debate."[119] The consensus achieved through the international rejection of the doctrine of racial supremacy is rare, Matsuda suggests, but this does not justify conflating it with other, relevantly different types of speech. This condition substantively distinguish racist hate speech from other forms of speech and as such, an explicit, content-based distinction is necessary. If the substantive difference is accepted, then it follows that a *sui generis* category is better suited to serve the interests of civil liberties by enabling the First Amendment to retain its force while also regulating racist hate speech. As Matsuda explains, "explicit content-based rejection of narrowly defined racist speech is more protective of civil liberties than

---

[118] *Idem,* 37.

[119] *Ibid.*

the competing-interests tests or the likely-to-incite-violence tests that can spill over to censor forms of political speech."[120] Introducing a separate category for hate speech subsequently minimizes the risks of undermining protection for other categories of speech.

Note that Matsuda's argument requires that the shared principle be accepted as unquestionably true, and it is only in virtue of this acceptance that the state need not worry about suppressing speech which might contain some truth. Furthermore, the aforementioned condition of shared, anti-racist principles sufficiently restrict the category of claims which the law ought to treat as certain, thereby softening the force of her position. Moreover, Matsuda's position allows individuals to engage with and evaluate the doctrine of racial supremacy, so long as such discussions do not additionally meet the conditions of targeting a historically oppressed group while also containing hateful messages of persecution and degradation. Matsuda's argument therefore allows her to grant Mill's arguments concerning the benefits of engaging with ideas both true and false. It merely prohibits the possibility that such ideas be used to target historically oppressed group through expressions which are hateful and persecutory. This argument is consequently *not* a counter-argument to Mill's original sub-arguments—it is instead an argument which can be taken in conjunction with his claims.

---

[120] *Idem,* 38.

Following the commitment to a narrow definition, Matsuda identifies three conditions, each of which is necessary for expressions to be deemed racist hate speech and thus, legally actionable. The conditions are as follows:[121]

(i) The message is of racial inferiority.
(ii) The message is directed against a historically oppressed group.
(iii) The message is persecutory, hateful, and degrading.

The first condition tracks explicit racism inherent in the expression, while the second condition highlights the continuous connection between racism and power. The third condition, on the other hand, tracks the degrading or dehumanizing nature of the speech. Taken in conjunction, these three conditions isolate a set of sufficiently harmful expressions which Matsuda takes to meet the legal threshold for restrictions based on likely degree of harm caused. It is also important to note the limited scope of such conditions, which exclude instances such as "arguing that a particular group is genetically superior in a context free of hatefulness and without the endorsement of persecution," satire, stereotyping, and "hateful verbal attacks upon dominant-group members by victims."[122] Despite the fact that these examples meet some of Matsuda's conditions, they nevertheless remain beyond the reach of legal action, however vitriolic the expression.

I.vi. Philosophical and Practical Concerns with Matsuda's Account

---

[121] *Idem,* 36.

[122] *Ibid.*

Matsuda's definition of racist hate speech is impressive in its ability to capture a restricted set of harmful expressions which meet a robust legal threshold for harm. However, deft as it may be, Matsuda's definition is not without its problems. Two such concerns, the problem of intra-group racist hate speech and the problem of reliance on shared, universal principles, are considered in detail.

A. Intra-Group Racist Hate Speech

While Matsuda's conditions filter an impressive number of difficult cases, at no point do they allow for a distinction between different kinds of speech act resting on differences in the origins of those acts. That is, they do not distinguish between kinds of speakers, but only consider to whom the message is directed. This inability to qualify the origin of the speech poses a problem in light of recent trends in the reclamation of oppressive speech, which popularizes intra-group usage of historically oppressive slurs by the original targets. And yet, these deliberative language practices seem beyond the scope of Matsuda's model, without reference to a speaker.

This lack is additionally complicated by the continuation of internalized racism and colourism, which leads members of historically oppressed groups to express vitriol

towards other members of said groups.[123] Incidents in which members of a historically oppressed group express "persecutory, hateful, and degrading" messages of racial inferiority towards other members of said group are not identical to incidents in which members of a historically dominant group expresses similar messages. The complexity of internalized racism which instantiates itself as vitriol expressed towards other members said group tracks a complicated internal state wherein such individuals call for violence against both themselves and their loved ones. This complicated internal state is a relevant moral difference and one which the law ought to track, excluding (or at the very least, treading lightly) with respect to the possibility of intra-group racist speech. Such incidents should therefore be considered on their own terms, rather than grouped together with cases of historically dominant group members expressing messages of racial inferiority. While it is possible that Matsuda did not think such usage could be legally actionable and subsequently omitted them from her account, this seems careless in light of the detail paid to restricting the set of incidents captured by her conditions. Another possibility suggests that Matsuda might have taken for granted the fact that such a claim precluded the possibility of intra-group hate speech, given that she goes to great length to stress that "racist speech is particularly harmful because it is a mechanism of

---

[123] Internalized racism often presents itself in the form of colourism—that is, assumptions and remarks which identify darker skinned individuals as less worthy or valuable than their lighter non-white counterparts. The multi-million dollar cosmetic industry dedicated to skin-lightning products is further evidence of the prevalence of such internalized colourism. Further discussion of the relationship between anti-blackness/colourism and hate speech can be found in Richard Delgado's "Words That Wound: A Tort Action for Racial Insults, Epithets, and Name Calling" in the *Words That Wound* collection.

subordination, reinforcing a historical vertical relationship."[124] While also possible, it nevertheless seems careless on her part.

Providing that Matsuda does not offer any further guidance, I take the likeliest explanation for the lack of qualification to be the result of multiple group oppression. That is, the lack of qualification ensures that her definition does not rule out the possibility of members of one historically oppressed group harming members of *other* historically oppressed groups using racist hate speech. Given the prevalence that colourism plays between different groups, omitting a condition concerning the speaker allows Matsuda to capture racist hate speech, regardless of whether the source of such speech is a member of another historically oppressed group. White individuals are far from the only source of racism, and a definition of racist hate speech must also capture the racist speech expressed by other, non-white groups.[125] If it is the case, then she is right to not want to rule out such a possibility. Then again, qualifying the second condition would not require that the speech come from a particular group, but merely that the racist hate speech *not* stem from a member of said historically oppressed group under consideration. As such, Matsuda's second condition could be qualified as follows: "The message is directed against a historically oppressed group *and expressed by a member*

---

[124] *Ibid.*

[125] A common example of non-white racism can be found in the racially-charged relationship between Middle Eastern countries, North Africans countries, and East, West, and South African countries. Middle Eastern countries such as Lebanon and Syria are often carelessly racist towards North African countries such as Morocco and Algeria, and in turn, North African countries extend such racism to African countries southward, such as Kenya and Mozambique. Colourism, it should be noted, plays a tremendous role in motivating such tensions.

who cannot reasonably claim membership to said group." This phrasing importantly allows for the inclusion of biracial individuals as well as individuals who pass as members of another group while also explicitly preventing the possibility of intra-group racist hate speech. As such, Matsuda's second condition could be adjusted to included more complicated, intra-group speech which might otherwise unwittingly meet the conditions of racist hate speech.

B. The Problem of Shared Principles

Recall that Matsuda explicitly endorses the truth of at least one shared principle—namely, the rejection of racial supremacy or a racial hierarchy—in order to motivate her argument. It was noted earlier that her shared principle allows Matsuda to distinguish racist hate speech as substantively different from Marxist speech, and in doing so, circumvent Mill's arguments. However, the reliance on shared principles to justify a substantive difference between racist hate speech and Marxist speech leads to problematic consequences concerning other forms of vitriolic speech. More specifically, shared principles establish a standard for legal restriction attainable by no other type of hate speech. This standard is therefore not only unhelpful for restricting other kinds of hate speech, it actively complicates the possibility of such. Cognizant as I am of the fact that Matsuda explicitly sets aside other kinds of hate speech, I nevertheless take it to be a mistake to introduce an argument for hate speech restrictions which comes at the cost of

restricting all other types of hate speech. The reliance on shared principles complicates restrictions for other types of hate speech by entailing at least two worrying claims:

(i) Racist hate speech is distinctive such that it ought to distinguished both legally and socially from other forms of hate speech.
(ii) Non-racist forms of hate speech are insufficiently harmful to generate legal protection.

Is racist hate speech distinctive from other forms of hate speech such that it warrants distinctive legal restrictions? Matsuda's shared principles seem to suggest so, especially in light of its unique universal condemnation. Yet it is important to note that the existence of shared principles does not entail that racist hate speech is substantively different from other kinds of hate speech. It only suggests, strictly speaking, that the perception of the doctrine of racial supremacy has coalesced to be remarkably unanimous. It is also not to say that such unanimity is not possible for other kinds of hate speech, but merely that it has not occurred. Hence, the existence of shared principles does not entail that there is a necessary difference between the doctrine of racial supremacy and any other doctrine advocating a social hierarchy.

Furthermore, while I recognize that racist hate speech is particularly charged in the United States, there remain other kinds of hate speech used in the same context which are equally vitriolic. I therefore take the second entailment to be equally problematic. Racist hate speech is a terrible and pervasive form of hate speech, but it is far from unique in that regard. Undue emphasis on one kind of hate speech risks complicating legal action taken to quiet other forms. As such, a sufficient legal definition of hate

speech must both account for the plurality of the category while also recognizing the power of certain types of hate speech in particular contexts. Matsuda's emphasis on the latter condition leaves her vulnerable with regards to the former, which I take to be a strike against her model.

Concerning the second claim: While I agree with Matsuda that hate speech ought to be a *sui generis* category of expression, I disagree that the category should be localized to racist hate speech for the above mentioned reasons. But Matsuda's heavy reliance on shared principles entails that her argument cannot account for other types of hate speech, even in light of similar levels of harm. The shared principle renders permissible a content-based restriction which would otherwise constitute censorship by distinguishing such speech as substantively different. Any argument defending hate speech restrictions therefore requires another shared principle, or an alternative argument justifying the permissibility of such content-based restrictions. Given that one would be hard pressed to unearthed anything close to a universally accepted principle concerning the status of women, members of the LGBTQ+ community, or disabled individuals, a generalization of Matsuda's argument cannot rely on another shared principle.

Having said that, I take Matsuda's definition to be structurally sound and more importantly, adept at distinguishing permissible from impermissible incidents of hate speech. Such benefits should, I think, carry over to any further definition of hate speech. With this purpose in mind, I propose the following modifications to Matsuda's three identifying conditions of hate speech:

(i) The message is of *social* inferiority.
(ii) The message is directed against a historically oppressed group *and expressed by a member who cannot reasonably claim membership to said group.*
(iii) The message is persecutory, hateful, and degrading.

I would additionally add that in spite of the fact that I defend a broadening of the *sui generis* category to include other forms of hate speech, I am not advocating for relevant differences between kinds of hate speech to be overlooked. I am instead defending the similarities between different kinds of hate speech and the manner in which they harm their targets. While cognizant that such a position could be construed as essentializing or reductive of the nuances of different types of vulnerability, I set aside such differences in order to emphasize the similarities between different kinds of hate speech. I takes these similarities to be legally relevant in order to argue for hate speech restrictions, but only if they can be systematized in some way. These modifications capture harmful instances of racist *and* non-racist hate speech whose harm can be supported by a reasonable standard of empirical evidence such that it meets the legal threshold.[126] However, an alternative argument is still necessary in order to justify the permissibility of content-based restrictions on hate speech.

I.vii. An Alternative Argument

[126] By "reasonable standard of empirical evidence," I am referring here to a standard not unlike that used by the Supreme Court of Canada in the *R.v. Butler* decision. See the following: "While a direct link between obscenity and harm to society may be difficult, if not impossible, to establish, it is reasonable to presume that exposure to images can cause changes in attitudes and beliefs. The question is not whether there is conclusive proof of a causative link but whether Parliament had a reasonable basis for acting."

The notion of content-based restrictions is misguided. Framing the conversation in such terms entails a false dichotomy that the speech is either restricted on the grounds of its content, or it is not. This dichotomy leaves little room for a conception of hate speech which goes beyond the mere content of the expression. It also leaves little room for a robust discussion concerning how it is hate speech can harm its targets. But one can recognize the need for an "explicit content-based rejection of narrowly defined racist hate speech" while also drawing attention to the fact that hate speech is more than its vitriolic content. The need for regulation does not stem from the vitriol communicated but rather, what the hate speech is striving to attain *in communicating such vitriol.* The possibility that there is more at work in hate speech than mere content returns us to the question of speech and action. It is also, ironically enough, the line of argument which provides the justification for content-based restrictions. Hate speech is not merely a group of speech acts which conflict with shared principles, it is a substantively different category of speech acts which perform a normative function. This function, striving to weaken the normative status of its target group, is unique to all forms of hate speech and as such, the identifying condition which renders permissible content-based restrictions on hate speech.

As such, we will now turn from the question of hate speech's content to the question of hate speech's function—a question currently motivating much work in applied philosophy of language. Lynne Tirrell, one such philosopher, argues that every instance of communication bears with it additional functions concerning the licensing of

further inferences about both the conversation and the world. This framework is then used to clarify how language can affect normative changes in the world. The function of hate speech as I take it, the unjustifiable weakening of the normative status of its targets, becomes most easily apparent when we turn to hate speech's most extreme accomplishment, the most extravagant harm to be inflicted on a people.

We thus turn to genocide.

CHAPTER III

ON THE NORMATIVITY OF HATE SPEECH

Lynne Tirrell's 2012 "Genocidal Language Games" opens the final chapter of this thesis with an inferentialist account of language. Heavily inspired by Brandom and Sellars, Tirrell emphasizes a conception of language as functional, that is, as performing functions. This pragmatic account of language, I will argue, highlights the function of hate speech: striving to weaken the social-moral normative status of target groups. While hate speech can target any group on account of the degradation of an essential feature, the success of hate speech's function is grossly facilitated in cases where targets belong to historically vulnerable groups, given that hate speech parasitizes already familiar patterns of thinking and inferences concerning such groups.

The pragmatic account of hate speech, as a speech act striving to weaken the social-moral normative status of targets, therefore clarifies the central question as to how hate speech harms its targets. More specifically, it clarifies harm to be an indirect consequence of hate speech and not, as has been suggested previously, from the mere communication of a hateful proposition. That is, harm follows from the normative function of hate speech.

Furthermore, only two types of harm necessarily follow from hate speech: risk and intrinsic harm. Individual and collective harm, though highly likely, are not necessary consequences of hate speech. But I'll suggest that the necessity of risk and intrinsic harm, taken in conjunction with the likely occurrence of individual and collective harm, is sufficient to generate *some* societal response. While my argument does not strictly necessitate legal intervention *per se,* it necessitates some type of normative intervention —a condition I suggest the legal system is characteristically well-suited to fulfill.

I.i. Lynne Tirrell on Genocidal Language

First published in the 2012 collection, *Speech and Harm: Controversies Over Free Speech,* Lynne Tirrell's "Genocidal Language Games" tracks the role linguistic practices played in facilitating the 1994 Rwandan Genocide. Using an inferentialist role theory of meaning, Tirrell argues that hate speech directed towards the Tutsi people engendered action and thus, normalized violence against the Tutsis.

Tirrell's analysis relies on a conception of hate speech which starkly contrasts with hitherto considered accounts of language as mere propositional content. She also does not conceive of hate speech as the visible effects of the "hidden roots of racism"—or at the very least, she takes this metaphor to be insufficient. While she acknowledges the manner in which racism grounds hate speech, in tracking the social history of Rwanda between the Hutu and Tutsi people, Tirrell takes language to have been, in no small part,

a cause of the racism which enforced the caste system in that society.[127] The emphasis therefore shifts from hate speech as an effect of racism to hate speech as a phenomenon which *creates* racism, where it may not have previously existed.[128] If it is the case that hate speech is more than an effect of racism, then arguments relying on the claim that we should combat the "real" problem of racism, fail to properly identify that problem. Hate speech, according to Tirrell's conception, ought to be recognized as a central component of racism, rather than an extraneous effect, and this alternate conception is, in no small part, a result of her inferentialist framework. We therefore move to consider how inferentialist accounts of language conceive of the causal capacities of language.

A. On Language-Games

Inferentialist theories of meaning, largely popularized by Robert Brandom, defends communication as a network of inferences. Better known as "language-games," these networks of communication permit inferences to be licensed or blocked. Necessarily collaborative in nature, inferentialist accounts entail that language oftentimes "outstrip[s] our own mastery" and accomplishes more than we, as individuals, may intend. To use a term is therefore to commit oneself to justifying the use of the term,

---

[127] Lynne Tirrell, "Genocidal Language Games," in *Speech and Harm: Controversies Over Free Speech,* eds. Ishani Maitra and Mary Kate McGowan (Oxford: Oxford University Press, 2012), 177-187.

[128] Tirrell's discussion of the history of Rwanda as it pertains to relations between the Tutsi and the Hutu people, leaves open the question of whether racism existed between the groups prior to the intervention of Belgians and identity cards. She concludes that the relationship was unclear, and I follow Tirrell on the question.

including supplying the reference and defending its role in other speech acts which may follow.[129] In addition to the traditional inferentialist commitments, Tirrell introduces "expressive commitments" that speakers maintain to the *value* and *viability* of using terms in some fashion. A term's usage therefore entails three related commitments:

(iv) Explaining inferences which are licensed by one's use of such a term
(v) Explaining how the term can be extended
(vi) Explaining the value in using this particular term or the utility of any further inference licensed by the use of the term

The inferentialist model draws heavily on the work of Wilfrid Sellars, and Tirrell sustains such philosophical commitments in her article by drawing his categories of actions which might be taken in a language-game. In "Some Reflections on Language Games," Sellars identifies three categories of moves available to participants in a language-game which explain the relationship between language-games and the world:[130]

(i) Entrance moves
(ii) Language-Language moves
(iii) Exit moves

Entrance moves take participants from perceptions or experience in the world to a position within a language game. The movement is therefore from world to word, and Tirrell demonstrates the centrality of such moves through the example of naming a child. As she explains, "the first use, say of a newborn child's name, puts the child's name into use, into the game, as it were, and forges a connection between the child and what is said

[129] Tirrell, "Genocidal Language Games," 188.

[130] *Idem,* 207.

about her."[131] Entrance moves, and first use in particular, are central moves which establish the connection between a term and its object of reference in the world.

However, not all entrance moves are equivalent, and certain entrances into the language-game may carry powers or responsibilities that are lacking from others. Social position is one factor which might determine the power of one's entrance move or the responsibilities which arise as a consequence. Returning to the naming example, parents and guardians are generally accepted as having proper authority in naming children, while school teachers or strangers are not. This is due to their position respective to the child, and other powers and responsibilities which follow. This qualification concerning power and responsibilities therefore emphasizes power relations underpinning the connection between words and the world, the obvious implication being that some connections are less benign than others.

Following entrance moves, language-language moves become possible. These moves constitute all speech acts licensed by a prior language-language move or a language entrance move, and they are "often based on approved patterns" or inferences which have been previously endorsed.[132] Examples of accepted patterns of inferences for the proposition "X is a dog" include questions concerning the dog's breed, his coat colour, and so on. Alternative accepted inferences, depending on known background information, could also include X's treatment of women, X's friendly personality, and so

---

[131] *Idem,* 209.

[132] *Idem,* 210.

forth. Language-language moves therefore expand the inferential capacity of the terms used while also enforcing the original connection between the term and the reference in the world.

Finally, language exit moves concern inferences which take individuals beyond the language game. This category captures behaviours outside the language game enabled by the language game.[133] "[R]eal-life language games," Tirrell notes, "are integrated into ways of life, and so actions within the game result in changed permissions governing behaviours beyond the game."[134] A doctor's prescription is Tirrell's choice example, as the prescription permits the actions so described. Language games should therefore be recognized to include this function—they permit a set of actions in the world which follow from the game's licensed inferences.

B. On Deeply Derogatory Terms and Hate Speech

With this understanding of inferentialist accounts of language, we now turn to Tirrell's analysis of language's role in the Rwandan genocide. She introduces the category of "deeply derogatory terms," meant to capture terms such as *inyenki* (cockroach) and *inzoka* (snake) as well as more general, hateful terms. The extent to which this category overlaps with hate speech is unclear, however deeply derogatory terms seems to be a broader category than hate speech. As such, Tirrell's account will be

[133] *Idem,* 210-211.

[134] *Idem,* 210.

taken in conjunction with Matsuda's account in order to provide a legal definition of hate speech which is not overly broad and respects the boundaries of the right to freedom of expression. My account is therefore not meant to capture all deeply derogatory terms, but only those deeply derogatory terms which also meet the conditions of hate speech.

Assuming an inferentialist framework, then, we must consider how such a framework accounts for deeply derogatory terms, identified by a set of five necessary conditions. The conditions are as follow:[135]

(i) Insider/Outsider Function
(ii) Essentialism
(iii) Social Embeddedness Condition
(iv) Functional Variation
(v) Action-Engendering

The insider/outsider function of deeply derogatory terms distinguishes target groups from non-target groups. Tirrell tells us that such terms "mark members of an out-group (as out), and in so doing, they also mark the in-group as un-marked by the term."[136] Built into the insider/outsider function is a necessary hierarchy evaluating the "outsider" group to be less socially valuable than the "insider" group, thereby justifying the initial distinction.

Importantly related to the insider/outsider function is essentialism, Tirrell's second condition, which enforces the aforementioned hierarchy by highlighting the permanence of said "inferior" characteristic. As she explains,

---

[135] *Idem,*190-193.

[136] *Idem,* 190.

> "Derogatory terms used in propaganda usually both presuppose and convey that there is an essential difference between the groups in question. Essentialism fuels fear, generates hate, and purports to justify differential treatment. This condition does not require that essentialism be true, only that it be presumed."

Essentialism is therefore the key to justifying sustained differential treatment, insofar as it prevents the possibility of targets ridding themselves of their "outsider" characteristic. The permanence of essential characteristics also distinguishes insulting or offensive terms from those which are deeply derogatory. While insults wield substantial power, their impermanence entails that "the term is a critique, not an assignment of a basic ontological status."[137] Essentialism, on the other hand, forecloses the possibility of the target riding themselves of the "inferior" characteristic, thereby entailing that targets permanently deserve differential (mis)treatment.

This assignment of inferior ontological status is facilitated by socio-historical circumstances, according to Tirrell's next condition. The social embeddedness condition identifies the power of deeply derogatory terms as originating in historical and social context wherein target groups have been previously devalued or dehumanized. This force, lacking in insults such as "jerk," stems from "networks of oppression and discrimination, with the weight of history and social censure."[138] Deeply derogatory terms are therefore weaponized in light of their embeddedness in unjust social, economic and political practices.

---

[137] *Idem,* 191.

[138] *Idem,* 192.

Tirrell's fourth condition, the functional variation feature, concerns the multiple functions of deeply derogatory terms. As with all language, deeply derogatory terms serve multiple functions. "[U]nderstanding language," Tirrell explains, "requires us to see the multitude of uses to which we put our words and to resist reducing these functions to one."[139] Reducing the multiple functions of language to a single function prevents proper awareness as to the full capacities of such terms. In the case of deeply derogatory terms, spewing vitriol is only one such function, among rationalizing cruelty and justifying violence.

While Tirrell is right to highlight the importance of recognizing language's functional variation, it does not follow from the mere existence of multiple functions that all functions ought to be treated equally. That is to say, notably harmful functions can and probably should be prioritized. One can sustain this priority while also recognizing the term's multiple functions.

The priority of harmful function leads to Tirrell's fifth and final condition identifying deeply derogatory terms, namely action-engenderment.[140] Through licensed patterns of inference in language-games, deeply derogatory terms permit differential treatment to the "outsider" group beyond the language-game. This treatment constitutes the language exit move of deeply derogatory terms, encompassing behaviours ranging

---

[139] *Idem,* 192.

[140] *Idem,* 193.

from spitting on "outsider" individuals to physical violence and, in Rwanda's case, effective genocide.

C. (Re)defining Hate Speech

Returning to the category of hate speech, Tirrell's conditions must now be considered in light of the narrower category of hate speech. All five of her conditions are pertinent to hate speech and ought to be included in hate speech. However, the fifth condition sheds light on a central difference between deeply derogatory terms and hate speech. In doing so, it reveals the function of hate speech: necessarily striving to weaken the social-moral status of targets.

The action-engendering capacity of deeply derogatory terms, Tirrell tells us, is not restricted to physical and psychological actions, but also includes normative actions. "Sometimes the action engendered is to assign a status-function," she explains, and the consequence of such inferior status assignment is violence or gross mistreatment.[141] While Tirrell suggest that status assignment might only occasionally be the action engendered by deeply derogatory terms, I hold the distinct function of hate speech to be striving to weaken the social-moral normative status of its target group.[142] In doing so, hate speech creates (or widens an already existing gap) between the target group's proper

---

[141] *Ibid.*

[142] It should be noted that even if Tirrell does take all deeply derogatory terms to assign status-functions of lesser value, then it simply entails that our categories overlap much more than anticipated. My claims do not contest her arguments, they merely strengthen them.

moral status and social-moral status. Proper moral status, innate to all humans, requires that they be treated as full and equal persons. Social-moral normative status, on the other hand, tracks the normative status of groups reflected by their treatment in society and depend on the geo-cultural context under consideration. Historically vulnerable groups, for example, maintain proper moral status as equal to others, but have been historically attributed weaker social-moral statuses, as degraded or inferior beings. In striving to weaken the social-moral normative status of its targets, hate speech therefore encourages morally impermissible treatment of targets.

While all speech acts which strive to weaken its target's social-moral status can be harmful, Tirrell's social embeddedness condition introduces an urgency otherwise lacking. This is due to the fact that historically weaker social-moral statuses *significantly increases the likelihood of hate speech's success in performing its normative function.* Hate speech grounded in unjust social, political, and economic practices are imbued with a force they would not otherwise maintain—and this is key. As a consequence of morally impermissible socio-historical treatment, certain terms are loaded with an additional force which significantly increases the likeliness of the success of hate speech.

The capacity of hate speech to weaken the normative status of targets becomes clearer if considered in light of the aforementioned language-game moves. In the case of historically vulnerable groups, it is significantly more successful at fulfilling its function as a result of the fact that *such hate speech does not begin with an entrance move.* Or rather, the entrance move has already occurred as a consequence of social embeddedness.

Social embeddedness in unjust practices bears previously accepted patterns of inference concerning the target groups, and those unjust patterns of inference subsist in the language game. By patterns of inference, I refer to "familiar" ways of thinking about target groups by non-members of such groups in the way of stereotypes or generalizations. These patterns of inference subsist beyond the expiration of the historical wrongs being committed in the larger, societal language-game. Slavery, the Holocaust, and the Canadian Residential Schools are obvious examples of wrongs which bear unjust patterns of inference that continue to this day, but restrictions on women's socio-political participation, the psychologized criminalization of queer identities, and the "merciful killing" or forced confinement of the disabled are also wrongs which sustain patterns of inference. Rationalized cruelty and careless violence against groups serve as a heuristic which tracks previously unjust social-moral normative status and the likely lingering presence of unjust patterns of inference.

The subsistence of such patterns enables current hate speech, when communicated, relieve the entrance requirement for the speech and instantiate directly as a language-language move. That is to say, the reference between the "persecutory, hateful, and degrading" speech act and the target group is inordinately forceful from the moment such speech acts are communicated because they rely on unjust patterns of inference taken for granted by a significant (and powerful) number of participants in the societal language game.

Moving away from the significant case of historically vulnerable groups, we now turn to the procedure as to how hate speech accomplishes such a task. Recall that chapter two featured a brief argument concerning the logical form of hate speech and the practical impossibility of verifying such speech acts to be true. Drawing on Tirrell's choice example, she notes that universalization was a key inference in facilitating the violent exit moves of the Rwandan genocide. As she explains,

> "The application of '*inyenzi*' spread beyond the invading militia by an extension of the term to all who share their ethnicity. This is a simple, but common, logical error: 'All *inyenzi inkotanyi* are Tutsi, therefore all Tutsi are *inyenzi*'."[143]

Hate speech necessarily translates to the form "All X are Y", where X refers to some target group marked by an essential characteristic and Y refers to some hateful or degrading term which signals social inferiority. This universal inference is supported by lingering patterns of inference sustained in the societal language game, but might also be licensed by current social trends which themselves facilitate patterns of inference. As such, lingering patterns of inference is at least one, but not the only source which might facilitate the generalization from existential to universal qualification—from the claim "one or more members of group X is Y" to "all members of group X are Y."[144]

Considering Tirrell's conditions for hate speech, all of which I take to be necessary for hate speech, I return to the modification of Matsuda's three conditions for

---

[143] *Idem,* 212.

[144] While the question as to what came first, the unjust normative status or the historical wrongs committed against groups, is a tricky egg-or-chick situation, a determinate answer is unnecessary for the purposes of this discussion. My argument merely requires a relationship between previous unjust status and historical wrongs, not on a particular orientation of this relationship.

hate speech at the end of Chapter II in order to introduce one final modification. My previous definition took the following form:

(i) The message is of *social* inferiority.
(ii) The message is directed against a historically oppressed group and *expressed by a member who cannot reasonably claim membership to said group.*
(iii) The message is persecutory, hateful, and degrading.

Of Tirrell's five conditions, the Insider/Outside Function, and the Action-Engendering Condition are explicitly captured by the following modified definition. Given that this account is primarily concerned with the harmful function of hate speech, concerns over the functional variation associated with hate speech are technically pre-empted.[145] However, the Essentialism Condition, in justifying differential treatment through the ontological permanence of the less valuable characteristic, remains to be captured. This permanence, it should be noted, was implicit in Matsuda's original, explicitly race-based conditions, yet was lost in my generalization.

However, this account suggests that hate speech cannot be weaponized against historically hegemonic groups. The answer to this question relies on the singularity of hate speech's momentum. If we take the historical wrongs and patterns of inference to be the unique source of hate speech's capacity to perform its function, then it follows that hate speech lacking the social embeddedness condition cannot harm in an identical fashion. But as mentioned previously, social embeddedness need not necessarily refer to historical social embeddedness. A trend or momentum of hate speech targeting

[145] While I recognize that this claim might be contentious, I am deliberately setting aside alternate functions of hate speech for the sake of discussing the function of harm.

historically hegemonic groups might facilitate the success of its normative function—there is no reason to think this possibility cannot obtain. Insofar as there is no argument precluding the possibility that other sources might infuse hate speech with an inordinate force, there is no reason to take hate speech against historically hegemonic groups as benign on account that they have been historically hegemonic. In other words, a historically hegemonic group is not protected from the normative force of hate speech in virtue of historical dominance.

Moreover, demonstrating a substantive difference between hate speech targeting historically vulnerable minorities and hateful speech acts against historically hegemonic groups does not necessitate that the law ought to distinguish between such groups—that is, the existence of a substantive difference is insufficient, in and of itself, to warrant legal definitions which adhere to such a difference. Given the necessity of generality of the law, the strength of the arguments for general legal provisions might nevertheless outweigh arguments defending a substantive distinction. David O. Brink offers such a position in "Millian Principles, Freedom of Expression, and Hate Speech", wherein he defends a general legal provision which, when applied, would recognize the urgent force of hate speech targeting historically vulnerable groups.[146]

[146] David O. Brink, "Millian Principles, Freedom of Expression, and Hate Speech," *Legal Theory* 7, no. 2 (June 2001): 147.

As such, I follow Brink and rescind Matsuda's condition restricting hate speech to historically oppressed groups. My identifying characteristics of hate speech are therefore as follows:

(i) The message is of social inferiority premised on an essential characteristic.
(ii) The message is expressed by a member who cannot reasonably claim membership to said group.
(iii) The message is persecutory, hateful, and degrading.

These conditions, as with previous instantiations, are all individually necessary for some speech act to be properly categorized as hate speech. This is due to the function of hate speech. Striving to weaken the social-moral normative status of target groups requires that these three conditions be met. Like Matsuda, my characteristics capture only a limited number of speech acts while leaving the vast majority of speech acts undisturbed in order to respect the right to freedom of expression. As such, arguing for the social inferiority of a group premised on an essential characteristic, though morally problematic, is not legally culpable on this account. Academic discussions which compare competencies based on genders, races, and other essential features consequently remain beyond the scope of regulation, so long as the other conditions are not fulfilled. Furthermore, satire, cartoons, and other artistic mediums which are not persecutory, hateful, and degrading are also free from legal regulation.[147] This account therefore prioritizes freedom of expression while also recognizing the function of hate speech.

[147] Indirect speech acts, as exemplified by art, constitute a distinct line of argument which is beyond the scope of this chapter. However, any concept of hate speech must be capable of recognizing the nuances of artistic speech and the complexities of indirect communication.

II.i. The Harms of Hate Speech

Hate speech is therefore among the direct cause(s) of weakened social-moral normative status-assignment for target groups. What exactly does this entail for questions concerning harm, and how can this account answer questions concerning legal regulation? Firstly, and perhaps most importantly, such a claim introduces a new variable to arguments defending hate speech regulation. In doing so, it alters the causal relationship between hate speech and harm. Hate speech is often characterized as causing harm in the following way:

Hate Speech —> Harm(s)

However, introducing social-moral normative status-assignment into the relation leads to an indirect causal model, which would resemble the following:

Hate Speech —> [Weakened Social-Moral Normative Status —> Harms]

According to the second model, hate speech still strives to *cause* something—it strives to weaken the normative status of target groups. Yet harm, usually taken to be a direct consequence of hate speech, becomes an indirect consequence. This seems to weaken the argument for legal regulation—after all, arguments defending hate speech regulation have hitherto been weakened by the inability to demonstrate the sufficiency of the harm to meet the legal threshold and thus, warrant state action. If I am claiming that hate speech does not directly cause harms, but only does so indirectly, have I not therefore offered an argument *against* legal regulation of hate speech?

Not quite. This conclusion only follows if one takes the "harm" of hate speech to be restricted to the direct physical and psychological consequences of hate speech—-which incidentally, I do not. Rather, my model has expanded the categories of harm to include at least four types of harm which may be distinguished:

(i) Individual harm
(ii) Collective harm
(iii) Intrinsic harm
(iv) Risk

Some clarification is necessary concerning these categories. Firstly, I take this list to be non-exhaustive, and as such, welcome additional categories in conjunction. Furthermore, these categories might overlap in interesting and important ways, but the quartered distinction highlights relevantly distinct dimensions of harm. Finally, these categories echo much of the literature on the harm of hate speech. This is deliberately so, given that I take much of the literature on the harm of hate speech to be accurate—hate speech causes a variety of harmful consequences. But these variations do not contradict one another, and my focus is outlining the manner in which such harms coexist as a direct consequence of weakened social-moral normative status.

A. Individual Harm

Individual harm concerns the physical and psychological harms faced by targets when confronted with hate speech, largely constituted by personal attacks. Such harms are instantiated through unjust or violent exit moves, which include physical assault and

psychological mistreatment, targeting particular individuals on account of their bearing some essential characteristic.[148]

However, it should be noted that individual harm does not necessarily follow from hate speech. This is due to the fact that there are individuals who, as members of target groups, might not have experienced personal attacks or been harmed in such ways. Some target group members might lack first-hand experience of hate speech, others might be of the constitution to be able to ignore hate speech, or certain luckier individuals may be protected from hate speech through mitigating factors, such as familial or personal wealth. Whatever the reason, individual harm is not a necessary consequence of hate speech, even if it remains a highly likely one.

B. Collective Harm

Like individual harm, collective harm is a highly likely, but not strictly necessary consequence of hate speech. Collective harm refers to harms which groups experience qua members of the target group. Like individual harms, such harms are also instantiated through exit moves that lead to unequal treatment of groups as a collective. However, unlike individual harm, collective harm need not have a particular target in mind, given that the harm is systemic. Collective harms concern all those who bear the essential characteristic, as in the case of unjust social policies or societal treatment, reduced

[148] I take psychological trauma to (arguably) be an exit move which follows from psychological mistreatment. As a direct consequence of the speech, it is a kind of self-perception permitted by the speech act.

political power, unequal job opportunities, and restricted access to public spaces. Additionally, such harms can be explicit policies or implicitly enforced social norms, as in the case of Chicago's mid-century restrictive covenants, which continues to implicitly enforce racial housing segregation.

C. Intrinsic Harm

Intrinsic harm, better known as constitutive harm, is not constituted by unjust consequences. Instead, such a harm directly follows from hate speech striving to weaken the social-moral normative status of its target. Relying heavily on J.L.Austin's speech act theory, first presented in *How to Do Things with Words,* Rae Langton and Ishani Maitra, among others, argue that speech acts constitute a harm merely in virtue of their utterance.[149] As Maitra notes in while describing the act of betting, "[m]y words just *constitute* the betting act; there is nothing further I have to do, besides uttering those words, to perform that act." (emphasis not mine)[150] While there has been some controversy as to the nature of constitutive harm and whether hate speech constitutes

[149] While a great many arguments have focused on the harm of subordination, especially in the context of pornography and women, constitutive arguments are not restricted to such harms.

[150] Ishani Maitra, "Subordinating Speech," in *Speech and Harm: Controversies Over Free Speech,* eds. Ishani Maitra and Mary Kate McGowan (Oxford: Oxford University Press, 2012), 98.

harm distinct from its consequences, the language of proper moral normative status and social-moral normative status clarifies what I take to constitute intrinsic harm. [151]

Hate speech may succeed or fail in its function to weaken the social-moral normative status of its target. However, a necessary condition of hate speech attempting such a function requires the misalignment of the target's proper moral and the target's social-moral normative status. That is, hate speech must distinguish the social-moral normative status from the target's proper moral status in order to make hate speech possible.

The necessary distinction between social-moral normative status and proper moral normative status is obviously present in cases of successful hate speech. However, this distinction is also present in unsuccessful instances of hate speech. One must minimally question the alignment of the two normative statuses, regardless of the success of the speech act. By questioning the necessity of the alignment between the proper moral normative status and the social-moral normative status, hate speech affirms the possibility that such alignment need not be the case. To recognize the possibility that a target group's social-moral normative status need not align with its proper moral status constitutes

[151] I am deliberately avoiding the language of locution, perlocution, and illocution, insofar as the details of Austin's model will complicate this brief discussion and takes us beyond the scope of this argument.

prima facie a moral harm.[152] Intrinsic harm is therefore a necessary consequence of hate speech's function.

D. Risk

The final category of harm, risk, draws heavily from Matthew Kramer's work in *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and Its Consequences*. "Just as a failed attempt to commit mayhem is typically evil," Kramer notes, "so too is an instance of extremely reckless conduct that does not actually result in the horrific consequences that have been knowingly hazarded."[153] Recklessness, or knowingly bringing about reckless risks to others, is a type of harm which Kramer takes to be wrong in and of itself. This is not to say that there is no ethical difference between unmaterialized risks and their materialized counterparts, but only that such a difference instantiates itself legally by the degree of punishment.[154]

Risk is therefore the fourth category of harm perpetrated by hate speech as a consequence of weakened normative status. Insofar as hate speech strives to weaken the normative status of all members of the target group on account of the essential

---

[152] Such questioning, it should be noted, also facilitates later questioning of the target's social-moral normative status, thereby enabling the possibility of success for later instances of hate speech.

[153] Matthew Kramer, *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and Its Consequences* (Oxford: Oxford University Press, 2012), 205.

[154] *Idem,* 205.

characteristic, all individuals bearing such characteristics face a risk which necessarily follows from hate speech.

II.ii. The Legal Threshold, Reconsidered

Having outlined four possible categories of harm which arise through hate speech's normative function, we are now better positioned to discuss whether the harms of hate speech properly meet the legal threshold to justify regulation.

As mentioned previously, I do not take individual harm to be a necessary consequence of hate speech, given differences in temperament and contextual mitigating factors. Yet it is individual harm which seems to bear the crux of the legislative burden. Recall that in previous discussion of "A Theory of Freedom of Expression," Scanlon's account of the harm of hate speech omitted any reference to collective harm and intrinsic harm, with some discussion of risk. That is, individual harm seems to be the primary category evaluated by Scanlon and others arguing that hate speech fails to meet the legal threshold of harm. This is not to say that other categories are ignored, but merely that other categories are devalued in light of the primacy of individual harm.

This decision, to prioritize a certain category of harm in order to justify legal regulations is a value choice unshared by defenders of hate speech legislation. Matsuda and Delgado, for example, highlight the importance of individual harm while drawing equal attention to the collective harm of hate speech. The significant disagreement over hate speech and its relationship to the right of freedom of expression seems to stem from

this value choice. If individual harm is the sole standard used by theorists to analyze whether hate speech meets the legal threshold of harm, but individual harm is not a necessary consequence of hate speech, then it is unsurprising that hate speech is deemed incapable of meeting said threshold.[155]

And yet, the importance of individual harm to the law must not be understated, and the fact that individual harm does not necessarily follow from hate speech might be an argument against regulation, in and of itself. While I am doubtful of the plausibility of this argument, the priority of individual harm in the law is worth emphasizing nonetheless.

Moving from individual harm to collective harm, then, we move from one type of justice to another. Individual harm concerns the injustice that a particular individual is being targeted on account of some essential characteristic, while collective harm concerns the injustice that a collective is being targeted on account of a shared essential characteristic, and the simplest form of collective harm is that of reduced opportunity. As with individual harm, collective harm is a highly likely, but not strictly necessary, consequence of hate speech. Striving to weaken a target group's normative status leads not only to an increased likeliness of violent exit moves, but also to inferior conceptions of target groups. Inferior conceptions of target groups, such as viewing X group as "less reliable" or "less trustworthy" reduces social opportunities and growth for the collective.

---

[155] Interestingly, the Supreme Court of Canada cannot be counted among such theorists who prioritize individual harm over collective or social harm. Decisions in which the collective harm of hate speech has been recognized include *R.v. Keegstra* and *R.v. Butler*.

As such, even in the unlikely possibility that no other collective harm occurs, a reduction of opportunities is a collective harm which is likely to follow from hate speech—further motivation for hate speech regulation.

Intrinsic harm, insofar as it is an injustice solely at the normative level, might prove the least convincing category of harm for the sake of legal regulation. Intrinsic harm, I suggest, is a category of harm which subsists even in the absence of wrongful consequences because it is a harm distinct from the success of the hateful speech act. Intrinsic harm captures the moral impermissibility of questioning the alignment between the proper moral normative status of target groups and their social-moral normative status. As such, I take this category of harm to be necessary insofar as it necessarily follows from the function of hate speech.

But it is the final category of risk which is the central harm of hate speech and as such, central to the possibility of hate speech regulation. If one accepts the claim that hate speech strives to weaken the social-moral normative status, then the nature of the risk which follows hate speech is necessary present, regardless of the speech act's success. In the case of historically vulnerable groups, this risk is amplified as a consequence of the inordinate likeliness of success. The risk, in the case of historically vulnerable groups, has previously materialized and as such, hate speech becomes the threat of such risks materializing once more.There are therefore two claims which I take to follow from my analysis of hate speech's risk:

(i) I echo Kramer in suggesting that a risk constitutes a harm, in and of itself, towards others, regardless of whether the risk materializes.
(ii) All hate speech instantiates a risk towards its target, and hate speech poses a significantly higher risk in cases of historically vulnerable groups.

The problem with hate speech, then, is not the harm. Or rather, the harm is an indirect consequence which makes urgent a more fundamental problem—the normative function which gives rise to at least two necessary dimensions of harm and two highly likely dimensions of harm. If this is right, however, there remains one (substantial) concern. If hate speech strictly causes normative changes, then it seems as though I have pre-empted the possibility of legal regulation prior to beginning an argument in its favour, given that evidence of normative changes might be difficult, if not outright impossible, to gather.

I would respond by first highlighting that this evidence-based concern is one any argument defending the harm of hate speech must address. Granting that evidence tracing the weakening of the social-moral normative status, in and of itself, would be next to impossible, does not entail that no evidence of the harm of hate speech is possible. The normative capacity of hate speech instantiate itself as individual and collective harms in the world, and such harms continue to play an important role in the legal conversation surrounding hate speech. However, such harms must *not* be confused as the direct consequences of hate speech. Setting aside the normative component of hate speech, an argument which takes individual and collective harm to be the direct consequence of hate speech leaves itself in an argumentative bind—in claiming the direct harm, it must demonstrate this direct relation or risk undermining its claim. Additionally,

counterexamples, in the form of target group members who do not experience such harms, threaten to significantly undermine the argument to a much higher degree, thereby facilitating the claim that the harm of hate speech is inconsistent and thus, insufficient to justify legal regulation.

In recognizing the normative function of hate speech, however, the importance of individual and collective harms can be sustained *without* the necessity of such harms. Individual harms and collective harms therefore serve as a heuristic, tracking weakened social-moral normative status in the face of hate speech. Collective harms, in particular, play a central role as a normative cue and obvious cues of this nature include financial struggles at an overwhelming rate, little to no visible representation in politics, restrictions to particular geographical locations or neighbourhoods, wide scale difficulty completing secondary education, and uncritical, common-place stereotypes in the public sphere. Such normative cues are largely context-dependent, in both degree and kind. Hate speech's normative function can therefore be traced by the presence of normative cues which might serve as evidence to establish a reasonable link in a court of law, without running into objections spurred by a necessity claim. As such, an argument concerning legal regulation is not pre-empted by my account and in fact, provides greater explanatory power concerning the supposed "inconsistencies" of hate speech's harm.

And yet, even if my normative account of the harms of hate speech is accepted, it does not follow that hate speech demands *legal regulation* per se. It entails only that hate speech requires some type of societal response. The final section will therefore consider

why legal regulation is characteristically suited to blocking the normative function of hate speech.

III.i. Inferential Blocks and the Law

In striving to weaken the normative status of target groups, hate speech warrants some societal response. The question left to decide is whether it warrants a response on the part of the law, and the inferentialist framework presented at the beginning of this chapter provides guidance as to how we might determine so. Recall that Tirrell noted two possible responses to an inference in a language-game: licensing and blocking. Explicit blocking, it should be recalled, was the only manner in which an inference could be prevented from continued use in the language-game. Otherwise, the licensing of inferences occurred when language-game participants tacitly or explicitly accepted inferential steps.

Blocking, or explicitly questioning the use of the term, is therefore the single measure through which hate speech's normative function might be prevented, insofar as blocking forbids licensing of the hateful speech act. In other words, if the speech act is questioned, the language-language move cannot continue forward as the speaker will be forced to defend the viability and the value of their speech act. While the extent to which such blocking will be successful is largely a consequence of contextual factors—that is, the extent to which such patterns of inference are questioned in the larger societal

language-game—the speech act will be stripped of its capacity to some degree in virtue of blocking efforts.

Furthermore, just as Tirrell noted that not all language entrances are equal, different blocking moves are equally varied—they carry different powers and responsibilities, depending on distinct capacities and contexts. Therefore, individuals and institutions might bear different degrees of responsibility to block unjust patterns of inferences as a consequence of their capacities and resources. With these considerations in mind, we now turn to the necessity of hate speech legislation.

The attractiveness of individual blocking cannot, I think, be overstated. Individual blocking is generally agreed upon as a potentially useful tactic to combat hate speech, and in a perfect world, it would be sufficient to block the harm of hate speech. Individual blocking permits theorists against hate speech regulation the possibility of acknowledging the harm of hate speech while also defending the broadest conception of freedom of expression. It is also recognized as a helpful blocking tool by theorists defending hate speech regulation as strengthening society's commitment against unjust treatment. The difference of opinion therefore stems from whether individual blocking is sufficient to block hate speech's normative capacity without involving legal measures.

However, any substantial role attributed to individual blocking is largely pre-empted by my model in light of the fact that I take hate speech to be a primarily normative act. That is, highlighting the normative function of hate speech entails that individual blocking, in and of itself, is rendered ineffective. This is due to the fact that

individual blocking is simply incapable of addressing the intrinsic harm and risk which follow from the normative function. It is additionally insufficient to address the collective harm of hate speech. This does not entail that individual blocking is without use, but merely that it is insufficient to wholly address the problem of hate speech.

The normative function of hate speech therefore requires blocking efforts with normative impact, at the very least, and this is a key condition which strengthens arguments defending the necessity of legal regulation. It is well within the purview of the ordinary powers of the legal system to restrict impermissible or unjust normative changes among its constituents, even between private citizens. Legal systems actively regulate normative changes which are deemed impermissible, as in the case of bigamy, and such regulations are often taken to be justifiable.[156]

Furthermore, the law's unparalleled access to those living under its domain renders it capable, in theory, of blocking hate speech explicitly, consistently, and unquestionably. While it might not always be successful in practice, its significant resources leave it well-equipped to fulfill its blocking potential. Moreover, the prioritization of consistency and longevity in the production of its laws substantially increases the possibility that hate speech laws will be both consistent and long-lasting.

Finally, the promulgation of hate speech regulation is necessary insofar as licensing occurs tacitly as well explicitly. Tacit licensing, on the part of language-game

[156] I am not, of course, suggesting that the discussion surrounding the illegality of bigamy is not complicated and perhaps, problematic. But I rely on bigamy only to suggest that legal regulation of permissible and impermissible marriages are commonplace, even if such regulations seem to infringe on central constitutional freedoms.

bystanders or participants who silently agree with the hateful speech act, goes unchecked in a system which relies on individual blocking. Such licensing subsequently presents itself elsewhere in later language-games or behaviours beyond, and as such, continues to (strive to) weaken the normative status of its targets. In comparison, the explicit promulgation of hate speech laws entails that both tacit licensing and explicit licensing can be blocked. This is on account of the fact that the law blocks such inferences through its relationship with each individual, rather than rely on the participation of said individuals in language games with others. Legal blocking therefore does not rely on participants explicitly licensing the hate speech in order to block the speech.

This is not to say that hate speech regulation is sufficient to strip hate speech of its normative function, but merely that it is very difficult to begin addressing the normative capacity of hate speech without such regulation. As such, I do not claim sufficiency with respect to hate speech regulation—far from it. But I am arguing that the type of societal response required as a response to hate speech necessitates some normative clout, and the legal system is a primary candidate for such a task.

CONCLUSION

One year following the 1992 *R.v. Butler* Canadian ruling, a 1993 United Nations report on Rwanda noted injurious propaganda, in conjunction with absence of the rule of law and the absence of any system for the protection of ethnic minorities, was a significant factor which "facilitated violations of the right to life."[157] The 1993 report is one of the countless documents recognizing the substantial role hate speech played in justifying violence against the Tutsi people, a conclusion found in the midst and aftermath of a massacre.

---

[157] "Nidiaye Report on Rwanda 1993," United Nations, Economic and Social Council, Commission on Human Rights, accessed August 31, 2018, http://www.preventgenocide.org/prevent/UNdocs/ndiaye1993.htm.

This thesis opened with an excerpt from a legal decision recognizing the rational link between impermissible changes in attitudes and behaviour, and criminalization. It is only fitting that it close with a reminder of the terrible consequences hate speech can bring about, if left unchecked. Arguments defending the democratic legitimacy of the state on grounds of the right to freedom of expression are as important as their staunchest defenders proclaim—yet the value of such arguments is not lessened by the coexistence of hate speech regulation. It was the possibility of dissent, we must recall, rather than the dissenting opinion itself, which constituted the grounds for democratic legitimacy for Mills and motivated the right to freedom of expression. Matsuda herself took this point, recognizing the rigorous demand that we continue to interrogate our intellectual positions. Few, if any, argue against the right to freedom of expression, and its broad reach must be sustained.

But, as we saw with Tirrell's work on the Rwandan genocide, language is far from mere propositional content, and there are urgent reasons for theorists to recognize this fact. The conditions of the world can alter with a single slur, as Rwandans know well, and a blind eye towards this fact allows the normative function of language to run rampant. By framing hate speech in terms of its normative capacity, I strove to move the conversation beyond the reduced questions of psychological offence into the territory of deliberate justification of violence. To speak of hate speech, we may therefore conclude, is to speak beyond vitriolic and incendiary speech. Hate speech is a call to action and to

ignore the social, collaborative nature of hate speech, the manner in which it degrades its target and justifies the impermissible, is to ignore the warning of impending violence.

However, this violence is far from inevitable, and the *Butler* decision makes this clear. The acknowledgement of the "rational link" between legal intervention and the protection of targets is an official recognition of the law's extraordinary blocking potential. If successful, legal blocking will bear no mark save for the continued flourishing of its people, historically vulnerable or otherwise. If unsuccessful, however, we need only turn to Rwanda to see the devastating potential of hate speech.

BIBLIOGRAPHY

Brink, David O. "Millian Principles, Freedom of Expression, and Hate Speech." *Legal Theory* 7, no. 2 (June 2001): 119-157.

Crenshaw, Kimberlè Williams. "Beyond Racism and Misogyny: Black Feminism and 2 Live Crew." In *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* by Kimberlè Williams Crenshaw, Richard Delgado, Charles R. Lawrence III, and Mari J. Matsuda, 111-132. Boulder, Colarado: Westview Press, 1993.

Delgado, Richard. "Words That Wound: A Tort Action for Racial Insults, Epithets, and Name Calling." In *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* by Kimberlè Williams Crenshaw, Richard Delgado, Charles R. Lawrence III, and Mari J. Matsuda, 89-110. Boulder, Colarado: Westview Press, 1993.

Kramer, Matthew. *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and Its Consequences.* Oxford: Oxford University Press, 2012.

Maitra, Ishani. "Subordinating Speech." In *Speech and Harm: Controversies Over Free Speech,* edited by Ishani Maitra and Mary Kate McGowan, 94-120. Oxford: Oxford University Press, 2012.

Matsuda, Mari J. "Public Response to Racist Speech: Considering the Victim's Story." In *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment,* by Kimberlè Williams Crenshaw, Richard Delgado, Charles R. Lawrence III, and Mari J.Matsuda, 17-51. Boulder, Colarado: Westview Press, 1993.

Kwame McKenzie, Kwame. "Racism and Health: Antiracism Is An Important Health Issue." *British Medical Journal* 326, no.7380 (January 2003): 65-66.

Mill, John Stuart. *On Liberty.* Kitchener, Ontario: Batoche Books Limited, 2001.

R.v. Butler, 1992, 1 SCR 452.

Scanlon, Thomas. "A Theory of Freedom of Expression." *Philosophy and Public Affairs* 1, no. 2 (Winter 1972): 204-226.

Sullivan, Shannon. "Inheriting Racist Disparities in Health: Epigenetics and the Transgenerational Effects of White Racism." *Critical Philosophy of Race* 1, no. 2 (2013):190-218.

Tirrell, Lynne. "Genocidal Language Games." In *Speech and Harm: Controversies Over Free Speech,* edited by Ishani Maitra and Mary Kate McGowan, 174-221. Oxford: Oxford University Press, 2012.

United Nations, Economic and Social Council, Commission on Human Rights. "Nidiaye Report on Rwanda 1993." Accessed August 31, 2018. http://www.preventgenocide.org/prevent/UNdocs/ndiaye1993.htm.