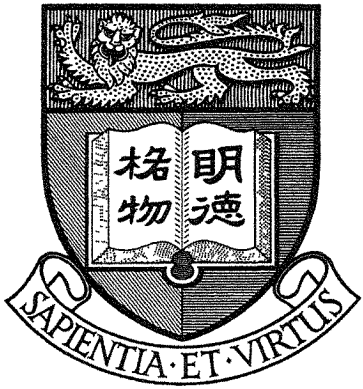
SHOULD "HATE SPEECH" BE OUTLAWED?

Wojciech Sadurski

THE UNIVERSITY OF HONG KONG

Faculty of Law
Law Working Paper Series
Paper No 11
December 1993



About the author

Wojciech Sadurski, Professor of Legal Philosophy at the University of Sydney, Department of Jurisprudence, is author of Giving Desert Its Due: Social Justice and Legal Theory (Reidel, 1985), Moral Pluralism and Legal Neutrality (Reidel, 1990) and numerous articles in legal theory and political philosophy. Fellow of the Australian Academy of Social Sciences and Editor of the Sydney Law Review, he is currently President of the Australian Society of Legal Philosophy. He presented this paper to the staff of the Faculty of Law, University of Hong Kong, on 6 October 1993.

SHOULD "HATE SPEECH" BE OUTLAWED?

Wojciech Sadurski

1. Introduction

Legal control of "hate speech" (i.e. of public vilification of certain groups on the basis of their race, gender, religion, sexual orientation, etc) poses a moral and philosophical dilemma of particular gravity to anyone committed to broadly understood principles of liberal and democratic legal order. This dilemma arises out of the need to balance and reconcile conflicting principles and concerns, each of which is valuable and important, and yet each of which must be subject to some preferential ordering in the specific setting of legal responses to "hate speech".

There are various plausible ways of characterising this conflict of values: freedom versus equality (which implies concern about a possible exclusion of victimised groups from the mainstream public discourse as a result of legal toleration of hate speech); formal equality versus substantive equality (the former, calling for an equal right to speak on matters of public concern regardless of the worth of one's views; the latter, implying concern about the dignity of those hurt by offensive speech); deference to majority opinions versus respect for minority interests (the former, relevant when the majority is unperturbed by the offensive speech of some extremists; the latter, demanding protection of the sensibilities of the most vulnerable groups), etc.

Regardless of the characterisation of the conflict of values in question, it is clear that the dilemma depicted is real and often dramatic. It is no wonder that there is a widespread feeling of frustration and perplexity concerning the approach of mainstream liberal theory to the problem, and also a certain angst amongst liberal theorists themselves who feel that it constitutes perhaps the most direct challenge to their established views about the limits of punitive state action. There is unfortunately a ring of truth to the statement that -- as Toni Massaro suggested -- it is precisely with respect to "hate speech" that liberalism's "vocabulary is least compelling, most paradoxical, and least responsive to real world conditions and actual human experience".[1]

The purpose of this paper is to take up this challenge and to attempt to elucidate the reasons for liberal inadequacy with respect to what I consider to be the key aspect of the harm produced by "hate speech".

It is clear that the characterization of "hate speech" as on balance harmful does not warrant, without more, a conclusion about legal prohibition. To allow speech to go unrestricted based on the condition that it is not on balance harmful would confine freedom of expression to an area so narrow as to be intolerable given the dominant values concerning freedom of speech. The regime of freedom of speech -- it is generally thought -- insulates public speech

from the use of harm calculus under which the government the government would be obliged to prohibit a speech act whenever it would be reasonable to expect that the resultant harm outweighed the harm of restraining people from speaking.

My principal concern in this paper is with two interrelated types of harm occasioned by hate speech, namely (1) "psychic harm", understood as an injury to the sensibility, dignity and self-respect of those who are targets of a given utterance, and (2) harm consisting in discrimination against disadvantaged, disempowered minorities. What makes this type of harm so special is that it occurs regardless of further consequences of a given pronouncement, for example, of whether it leads to an assault, violent response, or breakdown of social peace. To use a well-known formula borrowed from United States First Amendment jurisprudence, one might say that certain words "by their very utterance inflict injury".[2] In turn, what makes the second type of harm so severe is that it violates a fundamental precept of equality before the law.

This is, of course, not the only type of harm produced by "hate speech". One may think of at least the following other categories of harmful effects, irreducible to and distinct from "psychic harm":

1. Incitement to commit violence against the members of a target group;
2. Violent reaction by the victims of hate speech;
3. Degrading the general standards of civility and of public discourse in the community;
4. Hindering the education process (hate speech on the campuses), work relationship (verbal harassment in the workplace) etc.

For various reasons, these disparate categories of harm seem to me to be less problematic or less significant from the point of view of liberal theory than psychic harm. Harm consisting in the incitement to commit violence against certain groups (# 1) is captured by a general criminal-law prohibition against incitement to commit a crime. There seems to be no justification for applying different legal standards for incitement to violence against members of a group than for any other incitement to commit a crime. Keeping in mind a dictum that any advocacy of ideas may be seen as incitement, it is important to attack the evil of incitement with all the proper precautions against infringement of freedom of expression; such as requirements of mens rea, and of proximity between the incitement and the act of violence, and of "clear and present danger" etc.

Harm consisting in violence perpetrated by the targets of "hate speech", in reaction to that speech, seems also to be well

addressed by other rules of criminal law. The warning against establishing a "heckler's veto" (another useful idiom of American First Amendment jurisprudence) or against allowing a "hostile audience" to determine what can be said is worth considering. Unless certain circumstances occur in which it would be unreasonable to expect self-restraint from an offended person (in particular, if he or she is trapped in a situation of "captive audience" and subjected to face-to-face severe insult), the burden of responsibility should be on those actually initiating the chain of violent actions, regardless of the words uttered.

Harm to a community at large, consisting of the degradation of general standards of civility (# 3) seems to be less valid as a basis of legal restriction of offensive expressions. While real and severe, it is best handled by means other than legal restrictions -- on a broadly understood liberal theory, anyway. And finally specific harms which occur in particular types of environments, such as universities and enterprises (# 4), may warrant legal measures which are the least troubling from the standpoint of this paper. I am concerned here with the conflict of values in a civil society at large, where freedom of expression conflicts with protection of the sensibilities of minority groups, rather than with the balancing of values of autonomy and utility in specific settings, where a certain degree of deference to the judgement of decision-makers about what

is most conducive to an efficient accomplishment of their tasks seems to be well grounded.

2. Defining Group Vilification

"Racist speech", "hate speech", "racial vilification", "racist hate messages", "hate propaganda", "incitement to racial hatred" and so on are different names given in legal texts and in the literature to the types of group vilification with which I deal in this paper.

A useful way of indicating broad contours of what I mean by "hate speech" (or, more specifically, racial vilification) here this is to locate racial vilification on a continuum of types of offensive, dangerous or otherwise objectionable speech. At one extreme of this scale are the cases of speech that, even to the most committed libertarian, would uncontroversially warrant restriction. At the other extreme are the cases of speech that (even to those who support legal prohibition of group vilification) obviously should not be legally restricted.

We begin from the non-punishable end of the spectrum, where a consensus may be rather easily found against any legal prohibition or punishment of offensive speech. First, it will probably be widely agreed that statements of verifiable truth should be legally protected even if a particular group may find them offensive. Second, the context of a statement may suggest that neither was it prompted by wrongful motives (such as the desire to vilify a group)

nor that evil effects (such as offence caused) are likely to occur. Typically (though not uncontroversially) there will be cases of satire (but does that include racist jokes?), words uttered affectionately with no intention to hurt the hearer, a fair report of a public act, academic or scientific inquiry which may be distressing to some groups and yet is not motivated by hatred of those groups.

Third, no restriction is warranted when objections to an act of speech result from someone's excessive, unusual sensitivity. I do not want to get embroiled in the controversy about how easily measurable this standard is (it clearly is not), nor to worry about borderline cases. But commonsense suggests that there may always be instances of unusual sensitivity which should be disregarded.

Fourth, acts of speech which are private rather than public should escape the threat of legal prohibition. Just as in the preceding cases, this is more easily announced than applied. For example, the definition of a "public act" in the NSW Racial Vilification Act[3] is inescapably question-begging. It uses two independent criteria for "public act": it lists the forms of communication "to the public" which include "speaking, writing, printing" etc, and then it adds in a wholesale manner "any conduct [not covered by a previous description] ... observable by the public". It would perhaps be pedantic to ask "what is the public?" We may rather adopt a commonsensical approach that at least some types of speech are

undoubtedly "private" (such as inaudibly murmuring words to oneself).

While there may be other justifiable exemptions, these four categories seem to constitute the most important acts of potentially offensive speech that should remain legally permissible, and they will not be taken into account here as putatively prohibited instances of group vilification. Now to the other end of the spectrum: instances of the kind of speech that is uncontroversially non-protected. This end of the spectrum contains two categories: incitement and defamation.

A number of provisions, both in international and national legal instruments, refer to "incitement to violence" alongside racial vilification.[4] And yet, incitement to violence, constituting as it does incitement to commit unlawful acts, is beyond the focus of this article, since it does not entail any especially problematic issues from the point of view of freedom of speech. Now of course it may be argued that acts of racial vilification constitute a sort of incitement to violence or other unlawful acts. Indeed, it is frequently argued by the proponents of criminalization of group vilification that it is indistinguishable from incitement, or that it necessarily leads to incitement. But if indeed a given act of racial vilification amounts to the incitement of racially motivated violence, then there is no reason to resort to a separate crime of racial vilification.

The same applies to the commonly accepted indicia of defamation, as present in common law and various criminal codes. For all the differences between the laws of defamation in various legal systems, I will take it that prohibition of defamation is reasonably uncontroversial when applied to "individual" targets of defamation. One general principle in the English and American laws of defamation is that it will be punishable only when targeting identifiable individuals. This is traditionally expressed in the principle that a defamatory statement is not actionable unless published "of and concerning" the plaintiff.[5] Whether group vilification should also be punishable is a matter of separate substantive argument. It suffices to say that nothing is gained by extending the notion of individual defamation to group vilification, and it is clear that there are many who would find the former, but not the latter, a proper object of legal concern.

The upshot of the preceding remarks is that, for the purposes of this discussion, a statement will not be "racially vilifying" if it is verifiably true; or is made in a context that suggests neither a malicious intention nor a likely damaging effect; or is clearly made in private; or the resultant offence stems from unusual sensitivity of a member of a group. Furthermore, the concept of "racial vilification" as used here does not embrace statements that are

defamatory in the sense of individual defamation, or constitute incitement to unlawful acts against a particular group.

3. Psychic Injury

Liberal political philosophy is filled with arguments about the insufficiency of "mere offence" in warranting the prohibition of an offensive expression. Indeed, some liberals go a step further and claim that the fact that a given speech is offensive or distressing is all the more reason to protect it against restrictions. And yet, for all the attractiveness of the general principle, and for all the talk about protecting unpopular views against majority tyranny, there is a strange lack of correspondence between these proclamations and such personal experience of hurt and humiliation as are caused by racist slurs.

Liberals traditionally have had trouble grasping the severity of, and relating sympathetically to, the kind of psychic harm inflicted by group vilification. A possible reason for this is that they have usually associated psychic harms with images of majority offence occasioned by unorthodox or shocking minority behaviour. Within this framework, the authority of the state to protect people against harms to their sensibilities would effectively empower a majority to silence dissidents. But the pattern of the majority versus minority clash in the racial vilification laws is quite different. It is not a case of the majority trying to silence the minority but rather of a minority

trying to silence the intolerant majority, or trying to enlist the support of the majority in silencing a vicious minority. The silencing involved in enforcing anti-racial-vilification law is not the kind of silencing associated with majoritarian oppression. A group that seeks help through anti-racial-vilification laws is precisely the sort of group which has traditionally been seen by liberals as deserving special legal protection against possible majoritarian oppression: a powerless, subordinated and disadvantaged minority. So the whole pattern of minority protection versus majoritarian oppression is, in our subject-matter here, opposite to the one which prompts many liberals to raise the alarm against granting legal recognition to some people's aversion towards offensive speech.

It is important to note that the issue at hand is not just any offensiveness of a speech-act. We are concerned with a particular kind of harm to one's sensibility; that is, the harm resulting from vilification of one's own group. This is a different matter from a claim for protection from unwanted exposure to disgusting words and images. The difference is in the degree of the implication of one's own identity.

But the matter is more complicated than that. Consider the case of pornography: to some viewers it may be "merely offensive", but to those women who see it as expression of contempt for women, or part of an ideology that treats women as objects of sexual

exploitation, it is more than a "mere offence": it is an insult to themselves in a way which implicates their own identities. My point is not simply that we have a line-drawing problem between "mere offence" and offence that implicates the identity of a viewer/hearer. The problem is more serious: whether we classify a given offence as falling on one or the other side of the line depends ultimately upon our substantive assessment of the vilifying speech, and this has some disturbing consequences.

To reflect upon this, consider first the issue of "symbolic speech". According to some commentators, acts such as wearing a jacket with insulting or obscene words printed on it cannot be prohibited because, in contrast to group vilification, such acts do not harm anyone. In the words of a writer who is otherwise in favour of criminal liability for group vilification, symbolic offensive speech would escape punishment because "an essential characteristic of symbolic speech ... is that by definition it poses no serious harms to substantial public interests".[6] This is supposed to contrast with harm caused by "fighting words" and other categories of non-protected speech. But the distinction is ultimately in the eyes of the beholder: one may well imagine a situation in which harm to one's psychological well-being and self-respect, caused by an offensive or obscene sign, may be serious and (more importantly) may implicate one's own identity. Think of a committed nationalist who sees his or

her nation's flag depicted in a disparaging way on a T-shirt. The judgement that no serious harm to one's psyche is caused by a particular offensive sign hinges upon a judgement about the substantive value of this sign, and more importantly, about whether the viewer is justified in finding his or her own identity implicated by the sign.

At this point, it is useful to consider a distinction between "sensibility harm" experienced by the targets of group vilification, and "interest harm". The difference is said to be that in the case of "sensibility harms" the harm cannot be defined independently of the hearer's attitude towards the speaker's point of view, while in the case of "interest harms" it can.[7] Consistent with the currently influential "communitarian" and "civic republican" approaches, it is often claimed that "the law should recognize the sensibility harm experienced by the targets of group vilification, because to disregard it is to compromise the shared commitments that make freedom possible".[8] The immediate problem is of course that if we allow all "sensibility harms" to be protected by law, we shall end up restricting even the most valuable and justified speech, as long as the sensibility of a criticised person is harmed. What about civil rights demonstrations which may offend the sensibility of a racist?

To escape such counter-intuitive conclusions, the proponents of legal prevention of "sensibility harms" draw a further distinction

between "instrumental" and "constitutive" communities, the difference being that only in the latter ones members' identities are said to be directly affected by the statements related to the community as a whole.[9] Those same proponents of legal control of hate speech assure us that most communities are instrumental, and so do not warrant sensibility-harm based protection against group vilification. For instance, a white bigot will not be able to claim protection of his "sensibility harm" against civil-rightists, because the "communities" affected by anti-discrimination speech, such as racist groups, are not "constitutive".

The aim of the argument just summarized is to reconcile a prohibition of group vilification with the denial of such a protection to wrong groups. But the argument only pretends that it draws a neutral and clear line between "constitutive" and "instrumental" communities. In fact, I suspect that the operative line is between those communities of which we approve and those of which we do not. Consequently, the distinction has nothing to do with the seriousness of the implication of one's identity with the group of which one is a member. A Ku Klux Klan member may well be psychologically and morally affiliated with his organization to a higher degree than are many other people affiliated with their nations or their religions. His identity may be shaped by his Klan membership to a very high degree indeed, and yet we rightly deny

him protection from criticism of his organization. We do it not because we doubt the seriousness of his identification with the Klan but because we believe that its ideas are not worthy of protection. Hence, the distinction between "constitutive" and "instrumental" groups is just a proxy for a substantive judgement about those group identities that deserve protection and those that do not.

But if this is the case, then it prompts us to rethink our initial distinction between "mere offensiveness" and offensiveness that directly implicates one's own identity. This distinction is undermined unless we are prepared to say openly that it hinges upon our views about which group identities are worth protecting and which are not. If we <u>are</u> prepared to engage in such a value judgment about the worthiness of some sensibilities that deserve protection, then we must face the consequences of unrestrained majoritarianism. For example, we will disallow Muslim claims for special protection against blasphemy because we value their religion less than other religions. We think that their religious claims, even if genuine, are somehow less worthy in our society. And the same will have to be said to women who want protection from "sensibility harm" caused by pornography: their harm, even if genuine, is less recognized by the legislators as worthy of protection. I doubt if any legislator or legal scholar would be happy to accept this consequence.

The only alternative to drawing the distinction (between a mere offence and an identity-implicating offence) in this way is to abandon the distinction altogether, and to lower the level of legal protection for sensibility harm across the board. This, I believe, is a more candid and honest solution. The theory behind it may be summarized as follows: The severity of sensibility harms is in the eyes of beholders. Law does not draw the line between sensibility harms worthy of protection and those less deserving of legal concern. The degree of one's personal identification with his or her own wider group is not determined by the law: some people may have stronger identification with their church than with their nation, others -- vice versa. The law does not pass judgment on the quality of this identification. There is, therefore, no such thing as an agent-neutral, rigid distinction between "mere offence" and offence that implicates a person directly by insulting his or her group.

But the price to be paid for this legal abstention from judgements about offensiveness is that, in order to win legal protection, a claim for prohibition of group vilification must pass a very tough scrutiny of the speech-harm relationship. This is the inevitable trade-off: all complaints about vilification are considered genuine (i.e., as implicating one's own identity through vilification of the group), but since a discourse about groups is close to a public

discussion about political matters, in a self-governing democratic society it must be insulated against an ordinary calculus of harm.

This indicates the main difference, from the point of view of our analysis, between individual defamation laws and laws which punish for group libel. Even if both types of speech, covered by these laws, inflict similar injury upon the psychological well-being of the victim, individual defamation is usually so remote from discourse about public and political affairs that it may be measured by ordinary scrutiny without any risk to the value of self-government: the harm inflicted must simply be compared with the harm of restriction. But in the case of group vilification, the threshold of harm (to be demonstrated as produced by an act of speech) must be much higher because the speech in question is more relevant to the debate about public matters in a community.

This is not to say that racial vilification laws will necessarily fail the test, but that the test of harm must be much more stringent than the one applied, say, to individual defamation laws. The stringency lies, not merely in the requirement of a higher than usually demonstrated seriousness of harm, but also in the requirement of a proof that this kind of prohibition is the only way to avoid the harm. A state education policy and the enforcement of anti-discrimination laws spring immediately to mind in this context. If we consider the availability of such alternative ways of reducing

emotional harm, and also that harm is unlikely to follow from every speech-act that (under the racial vilification laws) would qualify for restriction, the harm of psychic and emotional injury is unlikely to overcome the hurdles of a strict scrutiny.

4. Hate Speech and Discrimination

Some of the more radical proponents of hate speech regulations have suggested that the exercise by racists of their right to make racist comments deprives their victim of their rights to free speech. If that were true, then prohibition of hate speech could be seen as a type of anti-discrimination measure. It would be tantamount to prohibition of "private discrimination" in the area of expression rights, analogous to outlawing "private discrimination" in the field of housing or employment.

The view that public presence of hate messages deprives their victims of some of their rights to free speech (the view that for brevity I will call a "silencing argument") can be understood in two ways, which raise different types of problems. It may be said, first, that unrestricted presence of hate messages actually prevents minority members from publicly expressing their views (or at least restricts their opportunities to do so),[10] or, secondly, it may be said that it devalues their views thus reducing the weight attached by public opinion to whatever minority members say.[11]

The first interpretation of the "silencing argument" seems to me to be plainly wrong. Literally speaking, no-one is prevented from, or restricted in one's capacity to, speak publicly by the very fact of someone else's hateful comments. But, of course, we need not be so literal: it may be claimed, for instance, that one has an important disincentive to speak if one is subject to vilifying statements by others. But then, conversely, it may be claimed also that one has all the more reason and motive to speak if one has to defend one's own group. A set of complex empirical factors affect this situation, including individual predispositions and, more importantly, how widespread the vilifying comments are. Arguably, in a society flooded by hate messages against a particular minority, there emerges a critical mass which may prevent one from making a case in one's defense. But the very presence, especially at a relatively low level of frequency, of hate messages, cannot be said to prevent one from speaking (although, as it was mentioned earlier, it deprives one of other goods). The situation is, of course, different when hate messages are expressed in a face-to-face, targeted, aggressive way, when they are more like assaults and less like communicative messages -- but this is precisely why my analysis here does not apply to such circumstances, and why I would have no problems with penalizing such behaviour.

The second version of the "silencing argument" raises a different problem. One cannot deny that, at times (and with respect to some members of the audience), hate messages may affect the credence given by the audience to views expressed by members of groups which are stigmatized by racist beliefs. There is nothing obvious about it, and one would hope that in modern societies the presence of racist views often has the opposite effect to the one desired by racists (i.e., it may strengthen the general respect accorded to the victims of vilification, and to what they have to say). But the devaluing effect certainly may, and does, occur.

The problem with this argument is, however, that it is indistinguishable from the argument that some views may be suppressed on the basis that they may cause some listeners to hold wrong views about certain groups. "Silencing" by means of devaluing what I have to say operates only through convincing other people (i.e. the audience) that they should treat me with less respect. In other words, the only way the "silencing" (in this second sense of the concept) operates is through persuading other people to hold certain beliefs about me, and about my group.

To prevent such "silencing" through suppression of wrong messages is anathema to a robust conception of freedom of speech. As David Strauss has convincingly shown in a recent argument, a generous conception of freedom of speech (as exemplified by the

First Amendment doctrine in the United States) is supported by what he calls a "persuasion principle": the principle which forbids the government from suppressing expression on the grounds that the speech may be effective in persuading people to hold certain views.[12] Strauss would explicitly deny the government the power to suppress "hate speech" on the ground that it may cause people to hold harmful views toward minorities.

The "persuasion principle" would not prevent the government from suppressing "fighting words" precisely because they are more like assault, and their persuasive effect is nil. But in the case of abstract, non-targeted views which express hate or contempt toward a certain group, the fact that these views may be persuasive is not a good reason to forbid their publication under a meaningful notion of free speech. The purported "silencing" which is at work here consists in the fact that due to the content of hate messages, members of the target group will become convinced that it is not worthwhile for them to make their points publicly and the substantial majority of the audience will become convinced that the members of the target group are not worth listening to. Now regardless of how likely it is that these two effects will follow as a result of hate messages (and, arguably, the former effect is likely to occur only if the latter occurs first), it is not the role of a government in a free country to prevent adult individuals from becoming convinced about whether it is

worthwhile to take part in a public debate, and about credibility or otherwise of various classes of speakers.

An extreme way of framing the second interpretation of the "silencing" argument is by saying that hate speech is a form of subordination of a target class. Suppression of hate speech would therefore be equivalent to governmental control over private coercion in a society. But to define racist speech, as Matsuda does, as "a mechanism of subordination, reinforcing a historical vertical relationship", and to say, furthermore, that racist speech is racism and therefore "is the structural subordination of a group based on an idea of racial inferiority"[13] is merely a facon de parler about expressions of subordination. To say that racial speech is identical with racial subordination is tantamount to a principle that speech which depicts (or even advocates) an odious social relationship should be suppressed on the weight of those same reasons which argue for a suppression of the relationship itself. No remotely robust system of free speech can survive under this principle.

It is important to understand why this principle is false: the reasons that the government has to enforce freedom of speech are not equivalent to the set of reasons that people have for speaking. Stanley Fish recently suggested that freedom of speech would be of important value only if speech were inconsequential, that is, if the reason people had for speaking was speaking for its own sake.[14]

Since, Fish argues, people speak not only to proffer propositions but mainly to achieve specific purposes, and those purposes may be defeated by some forms of speech, freedom of speech is chimerical. But this conclusion would be correct only if the rationale for freedom of speech was a mere aggregate of rationales of particular acts of speech, or, in other words, if there was a continuity between a legislative motive (for protecting citizens' rights to speak) and individual motives (for exercising these rights).

There are at least two reasons why this is not the case. First, governmental motives for decisions about limits on the right of expression are properly determined not only by an assessment of the value of purposes for individual expressions (in other words, of the value of the state of affairs produced by an aggregate of individual ideals) but also by what Ronald Dworkin calls "constitutive" reasons for freedom of speech: respect for individual autonomy, responsibility, etc.[15] Secondly, individual motives for advocating (or, conversely, for criticising) certain states of affairs are not always properly described as the purpose for the bringing about of that state of affairs. There may be various legitimate strategic reasons for advocacy: attempts to test public opinion, to strengthen an under-represented point of view, an "advocatus diaboli" argument, etc. The alternative, painted by Fish, that either speech is inconsequential or freedom of speech lacks foundation, is therefore

false. Hence, if reasons for freedom of speech are distinct from reasons for specific acts of speech, then neither are reasons for suppression of certain states of affairs tantamount to reasons for suppression of views which advocate those states of affairs.

5. Benefits of Racial Vilification

Now that the principal argument has been submitted, I may lift an underlying assumption about the overall harm of allowing "hate speech" made public, and consider whether there may be any good in allowing racial vilification to go unpunished. That is to say, is there any social good, other than the underlying value of strong protection of free speech, related to self-government in a democratic state?

First, one good of legality of such speech is that it might prevent a complacent attitude of society towards the existence of racism and racist attitudes. Racists are there, and it is better to let them air their views in the open rather than allow an illusion to grow that the problem has been solved because racist statements have been made illegal. Group vilification is a symptom, not a source, of deeper problems that give birth to hate and contempt by some people for other groups in society. By prohibiting public statements that vilify those groups we may slightly reduce the hurt to the feelings of their members, but at the same time we risk removing the issue of racism from the public agenda. The good of allowing group

vilification is that it helps maintain the visibility of a dramatic problem which is there anyway, regardless of the prohibition.

Second, legality of group vilification may, ironically perhaps, be valuable for the subordinated groups themselves. In a recent article, Kenneth Karst points out that unrestricted freedom of expression is a mixed blessing for the members of subordinated groups: on the one hand, they are victims of speech by members of the powerful groups in their society, but on the other hand "precisely because an important part of a group's subordination consists in silencing, their emancipation requires a generously defined freedom of expression, a freedom that overflows the shallow capacity of the model of civic deliberation".[16] I am not sure which benefit prevails in the calculus of costs and benefits for the subordinated and powerless groups, nor am I sure whether such a calculus can be made in abstract terms. And yet it seems to me that a broad regime of freedom of expression, which allows speech considered insulting, offensive and shocking to remain unpunished, may be at times useful to those groups who are alienated from the cultural and political mainstream. Conversely, it is likely that racial vilification laws may be invoked (even against the best intentions of their drafters) to silence the anti-establishment speech by disadvantaged groups, who might have to resort to shocking and "uncivil" expressions: to overcome the systemic bias of official channels of communications,

and to get their message across to the community at large (after all, "civility" is defined by a dominant culture which has all the conventional means of mass communication at its disposal). Remember the warning by Justice Black when he dissented from the Beauharnais decision which upheld an anti-hate-speech statute: "If there be minority groups who hail this holding as their victory, they might consider the possible relevancy of this ancient remark: 'Another such victory and I am undone'".[17]

Third, legal tolerance of group vilification has expressive value contained in the message it conveys to people about the nature of their society. To describe it in deliberately extreme and exaggerated terms, the message is as follows. In a liberal society claims to protection against insult and offense are viewed with utmost suspicion. Racial vilification approaches the outer limits of the rationale for tolerance, but this is precisely why the educational effect of this tolerance may be so powerful. Liberal society is opposed to a communitarian vision of legally protecting an individual's sense of identity with a wider group; under that vision, an attack on the group is viewed as an attack on an individual. In contrast, in a liberal society people are encouraged to distance themselves from their collective identities, to treat them as social roles rather than as ingredients of their own selves, and to put up with many dignitary injuries that other societies would have treated

much more seriously. To use a traditional distinction, a liberal society is much more of a Gesellschaft than a Gemeinschaft, or in a felicitous term recently coined by Meir Dan-Cohen, "a union of detached roles".[18] As contrasted to the ideal of "community", in such a union people are encouraged never fully to identify themselves with the roles they occupy, and their identity is never fully defined by the membership in groups. People are urged to be able to stand back from the roles they are playing, to adopt a critical attitude to them, and to perceive their own identity as transcending the sum of roles stemming from their involvement in various groups. This is one practical sense of the controversy about "thick" (i.e., context-bound) versus "thin" (or "unencumbered") conceptions of the self in a liberal society. It would not be easy to have "enough self-control to refrain from violent responses to odious words and doctrines"[19] if these "words and doctrines" were seen as destroying one's central aspects of self. But it can more easily be done if these insults are not allowed to reach your own identity: when they attack only some of your social roles, which are seen as extrinsic to your real self, you may easily "turn on [your] heels and leave the provocation behind".[20]

A determination of the degree to which your identity is constituted by your community involvements is, from this perspective, just another manner of speaking about how to react to

unpleasant or even odious utterances by others. A pluralistic, heterogeneous liberal society, must lift a number of traditional protections of one's psychic well-being, so as to maintain its pluralistic and cosmopolitan character. In a sense, a liberal society rejects the principle of honour as a good which one may protect through law; the idea of honour related to the community-defined individual is replaced by the central conception of an autonomous individual who may shape his or her social identities and attachments. Individual dignity is perceived more in the power of autonomous shaping of one's own social world rather than in the existence of a protected sphere of communal attachments which mould an individual self. Hence, this liberal insensitivity to many psychic harms is the price of a broadened scope for individual autonomy.

Many will think that the price is too high. Many will propose to draw the line between protection of individual dignity and protection of individual autonomy at another point, so that group vilification will fall on the prohibited side of legal restrictions. My main argument against penalizing group vilification does not hinge upon this expressive value of tolerating such speech. But this last point suggests that, whatever our substantive judgment about legality of these restrictions is, the importance of the problem is that it

prompts us to reflect upon the relations between the scope of legal prohibitions and the vision of the society we want to live in.

1 T.M. Massaro, "Equality and Freedom of Expression: The Hate Speech Dilemma", William and Mary Law Review 32 (1991) 211, 230.

2 Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942)

3 Anti-Discrimination (Racial Vilification) Amendment Act NSW 1989.

4 See, for instance, Article 20 of the International Covenant on Civil and Political Rights which declares that "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law".

5 See, e.g., Knupffer v. London Express Newspaper Ltd [1944] AC 116, 121 (HL).

6 Note, "Group Vilification Reconsidered", Yale Law Journal 89 (1979) 308, 322 n. 59 (emphasis in original).

7 Note, "A Communitarian Defense of Group Libel Laws", Harvard Law Review 101 (1988) 682, 692.

8 Id. at 692.

9 Id. at 693-94.

10 See Mari J. Matsuda, "Public Response to Racist Speech: Considering the Victim's Story", Michigan Law Review 87 (1989) 2320, 2337.

11 See Charles R. Lawrence III, "If He Hollers Let Him Go: Regulating Racist Speech on Campus", Duke Law Journal (1990) 431, 470-71.

12 David A. Strauss, "Persuasion, Autonomy, and Freedom of Expression", Columbia Law Review 91 (1991) 334.

13 Matsuda, op. cit., 2358.

14 Stanley Fish, "There's No Such Thing as free Speech and It's a Good Thing, Too", in Paul Berman, ed., Debating P.C.: The Controversy Over Political Correctness on College Campuses (New York: Dell, 1992) 231, 235-42.

15 Ronald Dworkin, "The Coming Battles over Free Speech", The New York Review of Books, 11 June 1992, 55, 57-61.

16 Kenneth L. Karst, "Boundaries and Reasons: Freedom of Expression and the Subordination of Groups", University of Illinois Law Review (1990) 95, 109.

17 Beauharnais v. Illinois, 343 U.S. 250, 275 (1952) (Black J, dissenting).

18 Meir Dan-Cohen, "Law, Community, and Communication", Duke Law Journal (1989) 1654, 1669.

19 Joel Feinberg, The Moral Limits of ethe Crminal Law, vol. 2: Offense to Others (Oxford University Press, 1985) 91.

20 Id.

Published by the Faculty of Law
The University of Hong Kong

1993

HK$15.00