Punishing Hatred and Prejudice

Heidi M. Hurd*
Michael S. Moore**

* Dean, and David C. Baum Professor of Law and Professor of Philosophy, University of Illinois College of Law
**Walgreen University Chair, Professor of Law, Professor of Philosophy, and Co-Director of the Program in Law and Philosophy, University of Illinois

This paper can be downloaded without charge from the Social Science Research Network Electronic Paper Collection:
http://papers.ssrn.com/abstract=472761
ABSTRACT

This article undertakes a detailed examination of the justifications advanced for the national and international rush to enact and apply hate and bias crime legislation as an answer to the tragically brutal expressions of racial animosity, bigotry, homophobia, and misogyny that continue to remind the Western World of its inability to protect its citizens from those who do not share its egalitarian ideals. In undertaking this project, we seek to synthesize and critically evaluate over a decade’s worth of scholarship on the wisdom of the enhanced penalties imposed by hate and bias crime legislation. We further seek to demonstrate that this literature, to date, has sadly failed to provide both an adequate moral justification and an acceptable doctrinal framework for this politically popular form of state action.

The article is divided into four parts, corresponding to the four principal rationales for hate/bias crime legislation that have advanced over the past decade or more. Part I considers the “wrongdoing thesis”—the claim that the harms perpetrated by offenders who are motivated by group-hatred or prejudice represent wrongs more serious than those perpetrated by defendants who commit the same offenses with different motivations. We work through an extensive catalogue of harms that are commonly claimed to be uniquely associated with hate- and bias-motivated crimes: elevated physical and psychic injuries to principal victims; wide-spread fear within the principal victim’s community; a diminished faith in the legal system and an associated instability within the larger social order; vigilante acts of retaliation by victims and their communities; the publication by such criminal acts of harmful messages corruptive of the moral order; and the associated, but independent harm of the state being complicit whenever it fails to express in law the moral outrage that such “statements” properly incite. As we demonstrate, even if social science ultimately vindicates the empirical claims made by those who propound the various wrongdoing theses, there are conceptual and moral problems that prevent these arguments from justifying the blanket sentence enhancements imposed by existing hate and bias crime legislation.

Part II takes up different versions of what we call the “expressivist thesis”—the thesis that the disrespect for communities expressed by acts of group-hatred and prejudice properly invite denunciation by the state in the form of elevated criminal penalties. As we argue, either the expressivist thesis is redundant with the wrongdoing thesis, and so invites the problems articulated in Part I; or it depends upon a free-standing expressivist theory of punishment, and is, for that reason, unsustainable.

Part III considers what we call the “culpability thesis”—the thesis that hate and prejudice constitute uniquely culpable mental states that justify penalties more severe than are meted out for other forms of viciousness. We demonstrate in this Part that if hate and bias are construed as culpability criteria, then hate/bias crimes are novel doctrinal inventions that are more at home within character-based theories of the criminal law that are best justified by political perfectionism, as opposed to act-based theories that are more in harmony with classic political liberalism.

Finally, Part IV takes up the “equality thesis”—the thesis that the enhanced penalties of hate/bias crimes properly function to achieve a more egalitarian distribution of the risk of crime within our society, because they deter the (further) victimization of groups of citizens who already bear a disproportionate amount of our society’s violence. As we demonstrate, each of
the various senses that can be ascribed to this claim render it either conceptually incoherent or morally indefensible, and as such, it fails to function as a promising alternative to the theories of hate/bias crime legislation that we examine in Parts I, II and III.
I. INTRODUCTION

This article undertakes a detailed examination of the justifications advanced for the national and international rush to enact and apply hate and bias crime legislation as an answer to the tragically brutal expressions of racial animosity, bigotry, homophobia, and misogyny that continue to remind the Western World of its inability to protect its citizens from those who do not share its egalitarian ideals. In undertaking this project, we seek to synthesize and critically evaluate over a decade’s worth of scholarship on the wisdom of hate and bias crime legislation. We further seek to demonstrate that this literature, to date, has sadly failed to provide both an adequate moral justification and an acceptable doctrinal framework for this politically popular form of state action.

Hate/bias crime legislation\(^1\) takes many forms and often specifies a number of different governmental objectives. For example, some so-called hate/bias crime statutes simply authorize

\(^{\dagger}\) Dean and David C. Baum Professor of Law, Professor of Philosophy, and Co-Director of the Program in Law and Philosophy, University of Illinois.

\(^*\) Walgreen University Chair, Professor of Law, Professor of Philosophy, and Co-Director of the Program in Law and Philosophy, University of Illinois.

\(^1\) Because as many statutes employ the language of bias as the language of hatred, and because it is one of the burdens of this Article to demonstrate that hatred and bias are conceptually and morally different, we shall adopt the cumbersome phrase hate/bias crimes o refer to all criminal
governmental agencies to gather information about the relative number of hate- and bias-motivated offenses committed within their jurisdiction within specified time periods.\textsuperscript{2} It is not with these enactments that we are concerned. Our interest is in the normative and doctrinal justifiability of legislation that contains sentencing enhancement provisions—provisions that authorize or require the enhancement of criminal penalties when already criminally-prohibited actions are performed as a result of offenders’ hatred or bias against their victims because of their victims’ race, ethnicity, religion, gender, disability, or sexual preference.\textsuperscript{3} This more narrow category of criminal legislation still permits wide variety in the manner in which the elements of hate/bias crimes have been defined and the degree to which penalties for already-criminalized conduct are increased upon proof of such elements. We shall not here provide any general summary of the varied language coined by different American states and foreign nations to define the necessary conditions for enhanced criminal penalties.\textsuperscript{4} Instead, we shall inquire generally into the different rationales which have been advanced to make moral sense of

\textsuperscript{2} Federal Act of early 1990’s, requiring reporting of incident statistics.

\textsuperscript{3} For our purposes, nothing turns on the specific characteristics included in this list. There has been much controversy, for example, over whether gender or sexual preference ought to count as “protected characteristics.” See, e.g., Frederick M. Lawrence, Punishing Hate, 15-20 (Harvard University Press, 1999); Steven Bennett Weisburd and Brian Levin, ‘On the Basis of Sex’: Recognizing Gender-Based Bias Crimes, 5 Stanford Law & Policy Rev. 21 (1994). Some has proposed that crimes motivated by a victim’s age or socio-economic status ought be thought of as hate/bias crimes. See, e.g., Michael Blake, __________, 20 Law and Philosophy (2001). Our analysis applies equally to generously and narrowly described categories of protected groups within hate/bias crime statutes.

\textsuperscript{4} For such summaries, see James B. Jacobs & Kimberly Potter, Hate Crimes: Criminal Law and Identity Politics, __-__ (Oxford University Press, 1998); Anthony M. Dillof, Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes, 91 NW U. L. Rev. 1015, 1020-23 (1997);
increasing criminal penalties for offenses committed as a result of racism, sexism, homophobia, and the like, taking up, when relevant, the doctrinal implications of such rationales and the fit between certain stated rationales and particular ways of doctrinally defining hate/bias crimes.

This article is divided into four principal parts, each devoted to one of the four rationales for hate/bias crime legislation that can be extracted from the considerable literature that has been amassed on this topic over the past decade or more. Part I considers what we call the “wrongdoing thesis”—the claim that offenders who commit crimes as a result of group-hatred or prejudice perpetrate greater wrongs upon their victims, or upon their victim’s communities, than do otherwise identical offenders. If, as is axiomatic to desert-based theories of punishment, punishment is properly attentive to the degree of wrongdoing inherent in a crime (and so is administered in greater degree to a murder than to a thief), and if a hate/bias crime is more wrongful than an otherwise-motivated crime, a hate/bias crime is appropriately punished with a harsher penalty.

In assessing the wrongdoing thesis, we shall initially treat wrongdoing as a function of the harms caused by an actor. On this view of wrongdoing a greater harm makes for a greater wrong. We shall accordingly work through an extensive catalogue of harms that are commonly claimed to be uniquely associated with hate- and bias-motivated crimes: elevated physical and psychic injury to principal victims; widespread fear within the principal victim’s community; a diminished faith in the legal system and an associated instability within the larger social order; vigilante acts of retaliation by victims and their communities; the publication by such criminal acts of harmful messages corruptive of the moral order; and the associated, but independent harm of the state being complicit whenever it fails to express in law the moral outrage that such “statements” properly incite. As we shall argue, even if social science ultimately vindicates the empirical claims made by those who propound the wrongdoing thesis from their armchairs (and it is nowhere close to so doing), such claims would not justify the blanket sentence enhancements of existing hate and bias crime legislation. For as we shall repeatedly demonstrate in considering each of the harms cited, among other problems, hate/bias crime legislation would then be using a defendant’s mental state as an indirect proxy for harms that are themselves directly, and more easily provable; and in so doing, it would violate the principle of desert upon which the wrongdoing thesis is thought to rest.

The second section of Part I assesses the greater wrongdoing thesis from a deontological perspective. Here we assess the possibility that hate/bias-motivated crimes constitute greater wrongs because violative of more stringent obligations of morality, even if no greater harms are caused thereby. We catalogue the variety of obligations that might make for greater wrongdoing here, but conclude that none of such obligations exist.
Part II takes up what we call the “expressivist thesis”—the thesis that acts of hate and prejudice express disrespect for members of the social community in ways that properly invite denunciation by the state in the form of elevated criminal penalties. This thesis takes two quite different forms. On one interpretation, this thesis is a special case of the previous wrongdoing thesis: in addition to being wrongful for other reasons, hate/bias crimes send offensive messages to groups who are entitled to equal respect, and they therefore constitute greater wrongs than do crimes that are otherwise-motivated (and that thus do not convey such messages). Construed in this manner, the expressivist thesis is at home with a desert-based theory of punishment; it simply seeks to identify a further reason to think that those who commit hate/bias crimes are guilty of greater wrongdoing, and are thus deserving of greater punishment than otherwise-motivated offenders. The second, more radical, interpretation of the expressivist thesis parts company from the idea that the criminal law should be (at least in part) devoted to achieving retributive justice. This more radical expressivist also disdains the traditional utilitarian accounts of the purpose of punishment. The radical expressivist appeals to an alternative theory of punishment—to the theory that punishment is justified if and only if it denounces conduct that is morally abhorrent to the majority of the community.

As we shall argue, the first interpretation of the expressivist theory threatens to be redundant with the claims of harm traditionally advanced by “wrongdoing theorists,” and so suffers from the same difficulties that we shall identify in Part I. The second interpretation of the expressivist justification of hate/bias crime legislation is only as strong as the expressivist theory of punishment upon which it rests. As we shall suggest, the best means of making sense of why it is obligatory upon the state to denounce hate/bias crimes ultimately collapses the second
interpretation of the expressivist theory into either a utilitarian theory (with all of the troubles that confront any utilitarian theory of the criminal law) or a retributivist theory (which then makes it redundant with the first interpretation of the expressivist theory).

Part III considers what we call the “culpability thesis”—the thesis that hate and prejudice constitute uniquely culpable mental states, and hence, that crimes committed with such mens rea are appropriately punished more than otherwise-motivated crimes. This thesis relies on the common notion that while punishment is in part a function of a defendant’s wrongdoing, it is also in part a function of a defendant’s culpability. Thus, just as a premeditated murder is properly punished more than a reckless homicide (even though both involve the infliction of the same harm, and hence, the same degree of wrongdoing), so, it is argued, hate/bias crimes are properly punished more than identical, but otherwise-motivated crimes.

As we shall argue, if hate and bias constitute new conditions of legal culpability, then the criminal law has been quite radically altered. For hatred and bias are unlike all other mens rea with which the criminal law has been traditionally concerned. Unlike intentions or goals, which can be abandoned by choice alone, hatred and bias are not directly responsive to the will. They are, rather, emotional states and enduring character traits that can, at best, be altered only indirectly, and only over time. For the criminal law to punish persons for bad emotions or bad character is thus for it to move from an act-centered theory of punishment to a character-centered theory, and so from a liberal agenda to a perfectionist one.

Part IV turns away from claims motivated by retributivist, mixed, or expressive theories of criminal justice to a thesis that presupposes that the criminal law is properly an instrument of distributive justice. According to what we call the “equality thesis,” the enhanced penalties of
hate/bias crimes properly function to achieve a more egalitarian distribution of the risk of crime within our society, because they deter the (further) victimization of groups of citizens who already bear a disproportionate amount of our society’s violence. As we shall demonstrate, each of the various senses that can be ascribed to this claim render it either conceptually incoherent or morally indefensible, and as such, it fails to function as a promising alternative to the theories of hate/bias crime legislation that we examine in Parts I, II and III.

II. THE WRONGDOING ANALYSIS

A. The Concept of Wrongdoing and the Distinction Between Wrongdoing and Culpability

“Wrongdoing” as we use it in this paper is a term of art, defined more by the concept’s role in organizing the structure of criminal law than by its ordinary connotations. Before coming to the various arguments concerning how or why bias-motivated crimes may be thought to consist of greater wrongdoing, we need to forestall certain predictable misunderstandings of what we shall say by elucidating the nature of wrongdoing itself. Legal wrongdoing standardly consists of: (1) a voluntary act, (2) causing, (3) some legally prohibited state of affairs. Each of

---

6 The criminal law does impose liability for omissions in some exceptional circumstances, but these exceptions are thought to prove the general rule articulated above. For a general discussion of the circumstances in which the criminal law imposes omission liability, and for an account of how omissions can be properly characterized as “wrongs,” see Michael S. Moore, Act and Crime: The Implications of the Philosophy of Action for the Criminal Law, 22-34, 48, 54-59, 267-278 (Oxford, 1993).
these three elements of legal wrongdoing, in turn, is rather elaborately defined. A voluntary act is traditionally defined as a willed (or volitionally-caused) bodily movement. This is taken to exclude involuntary bodily movements, such as reflex movements, hypoglycemic reactions, somnambulism, post-hypnotic movements, etc. It also excludes mere thoughts or wishes that do not issue in behavior. And it excludes the status that a person may occupy, such as “being drunk” or “being an addict”, since these are mere states that a person may be in, but not events that he brings about.  

A voluntary act causes some further event or state when it causes it both in fact and proximately. These ideas have elaborate, but contested definitions in criminal law. On the best view of causation, it is a real relation between events that is asymmetrical, intransitive, and highly discriminating in what is eligible to be a cause. Possession of these features is what makes causation so central to wrongdoing, for it is these features that make plausible the

---

7 Whether all legal wrongs prohibited by the criminal law have causal elements to them is a contested matter. Homicide obviously does, because the act prohibited (killing) is the same as causing death. Some argue that breaking and entering (burglary), taking (theft), and other “conduct crimes” possess no causal elements. This is certainly a mistake. Burglary clearly constitutes the causing of the state of affairs in which the defendant is on another’s property without permission, while theft clearly constitutes the causing of an owner’s dispossession of an item. See id. at 213-225.

8 See id. at 19-22.


identification of a wrongful act (such as killing) with a voluntary act causing some state of affairs (such as death).\textsuperscript{11}

The third element of legal wrongdoing exists because of the principle of legality honored by Anglo-American criminal law. The principle of legality holds that before punishment can be attached to the causing of some state of affairs, that state of affairs must be prohibited by a clear, non-contradictory, prospective, publicly available statute.\textsuperscript{12} At least this is true of prima facie wrongdoing—the legal wrongdoing that it is the prosecution’s burden to prove within our adversary system. Actual (and not merely prima facie) wrongdoing requires in addition the absence of any legally-recognized justifications, and such justifications are notoriously uncodified within Anglo-American criminal law.\textsuperscript{13}

We can now apply these preliminary considerations to distinguish wrongdoing from culpability.\textsuperscript{14} Wrongdoing (as defined above) has to do with actions. Culpability, by contrast, has to do with the actor, and specifically with the actor’s mental state. Mental states of desire, intention, and belief are familiar features of our psychology. They are the “Intentional” states with which we represent the world to ourselves. We represent the world as we believe that it is or will be, desire that it become, or intend that it be.

The criminal law uses such representational states to measure culpability. Roughly a defendant is more culpable if she intends to commit a wrong (as defined above, a voluntary act that in fact and proximately causes a legally prohibited state of affairs) than if she merely

\textsuperscript{11} Defense of the view that complex actions like killing are to be identified with voluntary acts-causing-deaths, may be found in Moore, Act and Crime, supra note 6, at ch. 8.

\textsuperscript{12} Id. ch. 9.

\textsuperscript{13} Moore, Placing Blame, supra note 5, at 74-75.

\textsuperscript{14} Id. at 403-405.
believes that her actions will cause that wrong to come about. And she is more culpable if she believes that she will do a wrong than if she should believe it, but fails to form such a belief because she unreasonably fails to make the requisite inferences about her behavior and its likely consequences. A defendant is prima facie culpable when she acts with one of these culpable mental states (intent, belief, or negligent oversight); she is actually culpable if she possesses one of these culpable mental states in circumstances in which she does not suffer from any impairments of capacity or limitations of opportunity which the criminal law recognizes as excuses.\footnote{Id. at chs. 12-13.}

One should see legal culpability as legal wrongdoing in the mind of the actor. This is because what a defendant must intend or predict in order to be culpable is just the legally-prohibited, causally-complex, voluntary action with which he is charged. The actor need not intend that his act be legally wrong, nor need he even believe that it is so; he must only represent his actions to himself under some description that the criminal law in fact deems wrongful.

Appreciating the distinction between wrongdoing and culpability allows one to see how non-idiomatic is the theoretical notion of wrongdoing we are using to organize this article. Consider two examples. In the first, actor A tries to murder his victim, V, by firing a pistol at V; but he misses. In the second, A knows V is behind a target but A shoots at the target anyway, not intending to hit V, but foreseeing that if he does what he intends to do—which is to hit the target—he may well miss the target and kill V. In point of fact, he hits the target and not V. In both cases, it is perfectly idiomatic English to say that A did something he should not have done, that he did the wrong thing, that he acted wrongfully. In both cases A could be prosecuted for
some crime (of attempt and of reckless endangerment, respectively). Yet in neither case is A
guilty of wrongdoing in the sense relevant here. For in neither case does A cause a legally
prohibited state of affairs (as a death is caused in a killing). When we impose punishment in
such cases, we are punishing culpability (the intent or belief by A that his acts will cause a
killing) in the absence of any wrongdoing.

Two other examples may further clarify this critical distinction. The first is a case of
“objective justification:” A intentionally kills his victim V just at the moment that V was about
to kill A, although A did not know this latter fact. A in fact has done nothing wrong as we are
using the term, since he like everyone else is entitled to match the deadly peril of a culpable
aggressor with deadly force of his own. But in using such force A is very culpable, since in his
own mind he was killing someone who he believed to be a non-aggressor. We may well punish
A for attempted murder, but when we do so we are punishing him for his culpability alone; not
for any wrongdoing.

The final example is that of “imperfect justification:” A believes that V is about to attack
A with deadly force so A kills V; in fact, V was not attacking A, although all appearances were
to the contrary. A is not guilty of anything, but not because he did not do anything wrong: he
did—he killed an innocent person. This was wrongful, albeit reasonable in light of the
circumstances as they appeared to A. What is lacking is any culpability on A’s part: in his own
mind A was not doing any act made wrongful by the criminal law. A is thus excused (lacking in
culpability), not justified (lacking in wrongdoing).

Since not all criminal law theoreticians draw the wrongdoing/ culpability line in this way,
it may be well to say why it should be drawn. Part of the justification for slicing the pie this
way lies in the greater systemization of the criminal law which it makes possible. Drawing the distinction as we have done neatly divides elements of liability between objective elements that have to do with the real world and subjective elements that have to do with the actor’s state of mind. The objective/subjective demarcation is not some arbitrary distinction that happens to fit the wrongdoing/culpability distinction as we have drawn it. Rather, there is good reason to map the elements of liability in a way that respects this divide. This is because of the conceptual priority of the objective over the subjective. In order to figure out what someone must intend or believe in order to be culpable, one must first figure out what it would be actually wrong to do; in order to figure out whether someone was epistemically justified in taking some known risk of some harm, one has to know when someone would be (objectively) justified in causing that harm. Even if one is inclined to blame persons for culpability alone, one needs to separately analyze the objective elements of what we call wrongdoing in order to know what culpability is.  

More importantly, there is a substantively moral reason for dividing up the elements of liability in this way. Doing so gets right the level of blameworthiness of actors like A in our four examples. If one subscribes to the notion of moral luck—the idea that what one causes adds to one’s overall blameworthiness—then one wants to distinguish failed attempts from successful ones, risks from causes, (objectively) justified acts from (objectively) unjustified ones, and acts that are merely excused from acts that are (objectively) justified. One must attend to the distinction between culpability and wrongdoing in order to honor the basic moral insight that

---

16 For a much more lengthy defense of the conceptual priority of wrongdoing over culpability, see Heidi M. Hurd, *What in the World is Wrong?*, 5 J. Contemp. Legal Issues 157-216 (1994).
blameworthiness depends not only on what is in the actor’s mind, but also on what he does in the real world.

The second distinction we need here, in addition to that between wrongdoing and culpability, is the distinction between two kinds of moral wrongdoing. On any plausible theory of the criminal law there will be a tight connection between moral wrongdoing and legal wrongdoing. Thus, any distinction between kinds of moral wrongdoing may well have an analogue in the criminal law.

As is well known, moral theory is sundered between “consequentialists” and “deontologists.”\(^\text{17}\) For a consequentialist, a right action is one that maximizes good consequences. It is such consequences, or states of affairs, that are bearers of intrinsic goodness or badness, and a human action is morally evaluated by the value of the states of affairs it produces. By contrast, for a deontologist, a right action is one that conforms to moral norms such as “thou shalt not kill.” That an action conforms to such norms is intrinsically good; there is no need to derive the goodness of such an action from the more basic goodness of the states of affairs that it produces.

Any plausible theory of morality should have room for both kinds of right action. No one can live by deontology alone,\(^\text{18}\) and a purely consequentialist morality is both too demanding and

\(^{17}\) See generally Hurd, *What in the World is Wrong?*, supra note 16; Moore, Placing Blame, *supra* note 5, at Ch. 17.

\(^{18}\) For more lengthy discussions of this claim, see Heidi M. Hurd, *The Deontology of Negligence*, 76 B.U. L. Rev. 249,253-54, 266-68; Hurd, *What in the World is Wrong?*, supra note 16, at pp. 196-99.
too lenient.\textsuperscript{19} Morality is best seen as governed by the general (consequentialist) obligation to maximize good states of affairs, subject to both agent-relative permissions not to do so in many circumstances and agent-relative obligations that are categorical in their force (and that thus prohibit their own violation in the name of good consequences).

On this eclectic view of morality there will be two kinds of moral wrongdoing. One consists of doing actions productive of bad consequences when morality contains no permissions to do so, and the other consists of doing an action categorically forbidden by morality. Torturing an innocent is a plausible example of deontological wrongdoing, while causing the death of another by negligently speeding is a plausible example of consequentialist wrongdoing.\textsuperscript{20}

Within a consequentialist account of wrongdoing the notion of harm is crucial. Harm is the common label for a bad state of affairs involving a person (and thus actions causing harm are consequentially wrongful). Harm is not an important notion for deontological wrongdoing, for the wrongness of an action is not derivative from some more basic badness of any state of affairs, harm included. This point is obscured because in fact the most serious deontological obligations are those that prohibit actions which cause serious harms to innocent persons. But not always. There are what Joel Feinberg once aptly called harmless wrongdoings,\textsuperscript{21} actions that are morally wrong even though they harm no one. Depending on one’s view of morality, “victimless immoralities” involving, say, sex or drugs, may be examples of harmless wrongdoing.

\textsuperscript{19} As too demanding, see Heidi Hurd, \textit{Liberty in Law}, 21 Law and Philosophy 385, 430-445 (2002); as too lenient, see Moore, Placing Blame, \textit{supra} note 5, at 680-684.

\textsuperscript{20} For a lengthy defense of the claim that negligence is consequentially, but not deontologically, wrongful, see Hurd, \textit{The Deontology of Negligence}, \textit{supra} note 18 at 249-72.

\textsuperscript{21} Joel Feinberg, Harmless Wrongdoing (Oxford 1986).
Yet even when the morally forbidden act does involve serious harm to another, the difference between the two kinds of wrongdoing survives. Death is an intrinsically bad state of affairs, yet one cannot capture the categorical force of morality’s deontological prohibition on murder by a consequentialist obligation against causing unnecessary deaths. What is wrong with murder, on the deontological view, is not that someone dies; rather, it is that the murderer has violated the categorical obligation not to kill intentionally.

The relevance of there being two kinds of moral wrongdoing is this: in proportioning punishment to the degree of wrong done, it is crucial to determine which kind of wrongdoing is involved. In the case of consequentialist wrongs, greater wrongdoing consists of causing greater harms; in the case of deontological wrongs, greater wrongdoing consists of violating more stringent moral norms. Such stringency for a deontologist is not cashed out in terms of harms. It is possible for there to be greater wrongdoing (if we are dealing with deontological wrongs) even if there is no greater harm caused. We will thus need to examine separately, as we do below, whether hate/bias-motivated crimes are worse because they produce greater harm, or whether they are worse even if they do not produce such harm.

B. Hate/Bias Crimes as a Source of Greater Harms (and Therefore of Greater Wrongs)

The literature that favors enhanced punishments for hate/bias crimes identifies four sorts of harms that hate/bias-motivated acts are thought to cause. Each of these harms is said to support the conclusion that bias-motivated criminals are guilty of greater wrongdoing, and hence, deserving of greater punishment.
1. Greater Physical Injury to the Victims of Bias Crimes

The first of the harms thought relevant by “harm theorists” about hate/bias crime legislation is that of greater physical injury. The claim is that bias crimes produce greater physical injury to their principal victims than do otherwise-motivated crimes, and that therefore hate/bias crime perpetrators deserve more punishment. Frederick Lawrence, for example, urges that “crimes committed with bias motivation are dramatically more likely to involve physical assaults than do crimes generally” and that “bias-motivated assaults are far more likely than other assaults to involve serious physical injury to the victim.”22 These two claims are quantified by Weisburd and Levin, who report that “bias crimes . . . are four times more likely to involve assaultive behavior than are crimes generally,” and that “bias-motivated assaults are . . . twice as likely to involve injury to the victim and four times as likely to require hospitalization.”23 Since, for Lawrence, “harm . . . lies at the heart of measuring the seriousness of a crime,”24 he concludes that the causing of greater injury merits greater punishment.25

Lawrence and his followers tender a puzzling claim here. One way to get at the peculiar nature of the claim is by imagining the following argument for enhanced penalties for crimes done by people with bad tempers: first, crimes by bad-tempered people are four times more likely to involve assaultive behavior than are crimes generally, and second, assaults done by bad-

---

22 Lawrence, Punishing Hate, supra note 3, at 39.
23 Weisburd and Levin, On the Basis of Sex, supra note 3, at 23.
24 Lawrence, Punishing Hate, supra note 3, at 58.
25 Id. at 60.
tempered people are twice as likely to involve injury to the victim and four times as likely to require hospitalization as compared to assaults generally. Therefore, whenever a bad-tempered person does something criminal, his punishment should be enhanced over the sentence received by normal-tempered people for the crime in question.

What is wrong with this argument should be apparent. The bad temper of the perpetrator is not itself a harm caused by the perpetrator’s action. So the argument above uses bad temper as a proxy for the greater harm of more serious physical injury. Yet there is no justification for using any such proxy. Proxies are almost always both over- and under-inclusive of the phenomena for which they are proxies: bad-tempered people do not always do crimes involving assaults, nor are their assaults always productive of greater physical injury, and normal-tempered people do sometimes assault others and their assaults are sometimes productive of quite serious injuries. Moreover, proxies involve averaging sentence severity across a class of offenders rather than individualizing sentences to each offender’s desert. While a rational penal code drafter might justifiably shoulder these two inevitable costs of using a proxy if, for some reason, proof of the fact of greater injury were typically impossible, here there is (and can be) no such showing. Whether a crime is an assault, and whether the assault produces serious physical injury, are provable in each individual case. Indeed, such facts are more easily proved than is the proxy fact of a defendant’s bad temper.

The only temptation that this imagined bad-temper argument can give rise to is one stemming from a confusion of culpability and wrongdoing. One might think that bad-tempered assaults are more culpable assaults, and on that basis one might attach greater blameworthiness to such assaults. But this would not be to use bad temper as a proxy for a victim’s greater harm.
To eliminate any unwanted confusion with culpability claims (which we will separately examine in Part III), one should view bad temper like red hair—a fact totally irrelevant to culpability but used for sentence enhancement solely because of its alleged power to serve as a proxy for greater harm to the victim.

The argument for enhancing punishments for hate/bias crimes based on the claim that they often involve greater violence and greater physical injury to their victims is fully analogous to the imagined argument based on a defendant’s bad temperedness. Once one sees clearly that one would be using a defendant’s hate/bias motivation as a proxy for a victim’s greater average harm—and not as an indication of the defendant’s culpability—it should be apparent that it provides no justification whatsoever for enhancing punishments. That biased individuals tend to commit more assaults than non-biased individuals is as relevant as the fact that the same can be said of bad-tempered individuals. In both cases the answer is that those who actually commit violent assaults—whatever their motive or their temper—should be punished for aggravated assault (as indeed they are under present law). And if the complaint is that the categories of aggravation specified by the law of aggravated assault are too coarse-grained to capture accurately the deserts of hate/bias-motivated assaulters, the answer should be that the criminal law should more finely individuate the levels of aggravation so as to make relevant to punishment more discretely-specified types of injuries. The answer certainly should not be that it substitute an even cruder, and less provable fact (the defendant’s motivation) for the provable fact of the victim’s aggravated injury.

2. Greater Psychological Trauma to the Principal Victims of Bias Crimes
A second common claim made on behalf of sentence-enhancements for hate/bias crimes is that such crimes cause greater psychological trauma to their immediate victims than do otherwise-motivated crimes. As Frederick Lawrence argues:

Bias crimes may also be distinguished from parallel crimes on the basis of their particular emotional and psychological impact on the victim. . . . A bias crime . . . attacks the victim not only physically but at the core of his identity. . . . This heightened sense of vulnerability caused by bias crimes is beyond that normally found in crime victims. . . . The victims of bias crimes thus tend to experience psychological symptoms such as depression or withdrawal, as well as anxiety, feelings of helplessness, and a profound sense of isolation.²⁶

Lawrence further maintains that because hate/bias crimes stigmatize their victims, and because “stigmatization has been shown to bring about humiliation, isolation, and self-hatred,”²⁷ hate/bias crimes also tend to cause their victims a loss of self-esteem. Other proponents of hate/bias legislation detail a rather longer list of traumas induced by hate/bias crimes.²⁸ They urge that victims of bias-induced violence experience greater feelings of vulnerability to future violence; fear verging on paranoia about strangers; feelings of self-blame; feelings of helplessness; depression; sadness; lack of trust in people; withdrawal from personal contact; sleep problems; excessive anger; suicidal thoughts; headaches. The claim in each case is that victims of hate/bias

²⁶ Lawrence, Punishing Hate, supra note 3, at 40.
²⁷ Id. at 41,
²⁸ See Weisburd and Levin, supra note 3, at 26.
crimes experience these traumas at a greater rate than victims of otherwise-motivated crimes, including otherwise-motivated crimes of violence such as robbery and rape.

The conclusion drawn from these empirical claims is that hate/bias crime offenders deserve more punishment because they are more blameworthy; they are more blameworthy because they have done greater wrong; they have done a greater wrong because they have (typically) caused more harm, namely, the psychological traumatization of victims (in the way just detailed). 29

In assessing this argument for hate/bias crime sentencing enhancements, the first item to question is whether the empirical claims in the argument are true. All crimes of violence tend to produce psychological trauma in their victims. As one study notes, to some extent, the predominant emotional responses of hate violence victims appears similar to those of victims of other types of personal crime. 30 The question is whether hate/bias-motivated crimes of violence produce significantly more such traumas than do otherwise-motivated but otherwise identical crimes of violence. Jacobs and Potter have recently reviewed some of the leading studies that claim to establish greater trauma for hate/bias crime victims. They conclude such studies are defective in not comparing the trauma of hate/bias crime victims with the trauma of comparable crime victims. 31 For without comparing ordinary assault victims with hate-motivated assault

29 Lawrence, Punishing Hate, supra note 3, at 63.
31 Jacobs and Potters, Hate Crimes, supra note 4, at 82-83.
victims (when the severity of the assaults and the injuries are equivalent), how can one isolate the traumatizing effects of hate and bias?

Sometimes those seeking to justify hate/bias crime legislation on this ground try to bolster the plausibility of their empirical claim in the following way. They piggyback the claim of heightened trauma on the claim that hate/bias crimes involve heightened violence. With more severe violence, it is of course plausible to suppose there is more severe psychic trauma. Yet a complete answer to this argument is provided by the fact that the law governing aggravated assaults adjusts punishments to the relative degree of violence and injury (including psychic trauma) perpetrated by aggressors. Part of what makes aggravated assaults worse than simple assaults is the greater psychic trauma accompanying more severe attacks and more severe physical injuries. That factor has already been taken into account in setting the punishments for the more aggravated assaults, so one would unjustifiably double count this factor by adding enhanced penalties to hate/bias motivated crimes because of the psychic traumas that attend more severe violence.

Despite these reservations about the empirical claim, let us suppose for the time that it were true: to some extent hate/bias crime victims experience greater trauma (fear, lowered self-esteem, anger, depression, etc.) than do victims of otherwise-motivated but otherwise comparable crimes. Such a supposition allows us to reach other problems with this argument for sentence-enhancements.

One such problem is the proxy problem we encountered before with respect to the claim that hate/bias crimes magnify physical injuries. Just as hate/bias motive of an actor is not itself a

32 E.g., Weisburd and Levin, supra note 3, at 25.
physical injury to a victim, so it is not itself a psychic trauma either. Those who justify hate
crime legislation on this basis thus are proposing that the defendant’s motive function as a proxy
for what really heightens the gravity of the wrong, namely, the victim’s psychic trauma.

But hate and bias are particularly bad proxies for a victim’s psychic trauma. It is not hate
and bias, as such, that hurt; rather, it is the perception of hate and bias that hurts. An assault that
is in fact motivated by hatred or bias, but which is not perceived by the victim as being so
motivated, can hardly be said to cause heightened psychic trauma. (Inversely, of course, an
assault that is not motivated by hatred or bias may nevertheless be perceived as being so
motivated, and so may indeed cause elevated psychic injury.)

An analogy is again perhaps helpful here. At one time criminal assault was defined as
attempted battery; an accused had to try to cause a harmful or offensive contact in order to be
guilty of criminal assault.\(^33\) Such a trying required proof that the accused intended to hit the
victim. Suppose that someone sought to justify this old version of the crime of assault by
arguing that what makes such an assault blameworthy is the fact that a victim’s apprehension of
contact is a genuine harm. That is, suppose that the gravamen of such a crime were viewed (as it
always was in torts) as a kind of psychic injury: the apprehension of imminent physical contact.
Could one in all seriousness propose that the intent of the offender to hit the victim would
function as a serviceable proxy for the apprehension of being hit of the victim? Surely not.
While an offender might try to hit a victim, the victim might be unaware of the near miss, and
thus have no apprehension of imminent contact. Or the victim could become apprehensive about
being hit when an offender intends only to scare, but not to hit, the victim. Surely a rational

\(^33\) See Rollin Perkins and Ronald Boyce, Criminal Law, Third Edit. 159 (Mineola, NY:
Foundation Press, 1982).
penal code that truly cared about the wrong done to a victim in causing apprehension of contact would define assault so as to make such apprehension an element to be proved in its own right, eliminating any proxy for this element.

At a minimum, the same should be done for hate/bias crime definitions. If the psychic trauma to victims of hate/bias crimes is thought to be the gravamen of the offense, then the law should not be concerned with the motive of the offender, but rather, with the perception of the defendant’s motive by the victim. Then hate/bias-motivated offenders who do not cause the psychic harms that, ex hypothesi, attend perceived hate/bias would not be subject to enhanced punishment for a wrong they did not commit; and offenders who are not hateful or biased in their motives, but who know or are reckless with regard to the fact that they will be taken to be hateful or biased\(^{34}\) should receive an enhanced sentence, by virtue of the greater wrong that they do in causing such a perception.

However, even if hate/bias crimes were redefined so as to make the victim’s perception of hate/bias the gravamen of the offense, there would still be an unacceptable proxy function operative in hate crime legislation. A victim’s perception of hate/bias in his attacker is still not the same thing as the experience of greater psychic trauma. Nor is the one universally accompanied by the other. A victim of a bias-motivated assault who perceives the bias in it may not suffer any greater psychic trauma than the typical victim of an otherwise-motivated assault;

\(^{34}\) Here we are simply honoring the presumption that any newly defined crime will contain a culpability requirement of at least common law malice, i.e., recklessness. In fact, of course, any newly specified crime of the sort we have here proposed in answer to the psychic trauma argument could raise or lower the culpability threshold. It could function to criminalize only the purposeful incitement of the perception of hate or bias, or it could extend to acts that simply negligently cause such a perception. It could even function as a strict liability offense, although it would then invite all of the well-rehearsed arguments against strict liability for non-regulatory offenses in criminal law.
additionally, one who suffers terribly in her mind after an assault need not suffer so because she perceived the assault to be biased in its motivation. Perceived bias, in other words, is still over- and under-inclusive as a proxy for severe psychic trauma.

So if one is truly concerned with adjusting penalties to the psychic traumas experienced by crime victims generally (and to the psychic traumas experienced by victims of hate/bias-motivated crimes in particular), the best thing to do would be to argue for a new form of aggravated assault, one that explicitly adds the element of the victim’s extreme emotional distress to the actus reus of the offense. Then when there is such trauma caused by an offender’s apparent hate or bias, there can be enhanced punishment of the offender, but not otherwise.

For our own part, we doubt that anyone truly believes that it is enhanced psychic trauma of victims that best justifies hate/bias crime legislation. It is not that we doubt that the severe psychic trauma suffered by particular victims of particular crimes should enhance the sentences of those offenders. It should. But such severe trauma is too ubiquitous to victims of violent crimes generally to be singled out for special emphasis through hate/bias crime legislation. Victims of hate/bias crimes should have what all crime victims should have: the right to present a victim impact statement at sentencing. Then their particular psychic trauma can be taken into account in measuring the wrongfulness of their offender’s deeds, no matter what aspect of the crime it may be that triggered their special harm. Targets of hate/bias crimes as a class simply are not different enough from other crime victims in their psychic injuries to justify enhancing their offenders’ penalties without specific proof of those injuries in individual cases.

35 The analogy would be to the tort of intentionally inflicting extreme emotional distress. See, e.g., W. Page Keeton et. al., Prosser and Keeton on Torts, sec. 12 (West, 1984).
3. Vicarious Injuries to Members of the Victim’s Larger Community

The psychological traumas suffered by the principal victim of a hate/bias-motivated crime are also said to be suffered by other members of the group(s) that share with the principal victim the characteristic(s) that prompted the defendant to target the principal victim. As Frederick Lawrence puts this claim:

Members of the targeted community of a bias crime experience that crime in a manner that has no equivalent in the public response to a parallel crime. The reaction of the target community not only goes beyond mere sympathy with the immediate bias crime victim, but exceeds empathy as well. Members of the target community of a bias crime perceive that crime as an attack on themselves directly and individually.\(^{36}\)

Members of the targeted community are thus said to stand in the shoes of the actual victim of a hate/bias crime, in that they too are caused to feel fear, vulnerability, intimidation, isolation, depression, etc. As Lawrence opines, “This additional harm of a personalized threat felt by persons other than the immediate victims of the bias crime differentiates a bias crime from a parallel crime and makes the former more harmful to society.”\(^{37}\)

\(^{36}\) Lawrence, Punishing Hate, supra note 3.

\(^{37}\) Id.
One’s first response to claims such as these may be to doubt the relevance of harms to persons other than crime victims in gauging the degree of wrong done by some offenders. After all, it is a serious wrong to murder another, no matter whether the one murdered was a much loved father of small children or a much despised old hermit. To say otherwise is to say that the wrongness of a defendant’s act turns on the popularity of his victim (that is, on how many family members, friends, and caring acquaintances he has). Those who find this implication unpalatable might thus be inclined to say that, as a moral matter, how others are affected by someone’s misfortune at being a crime victim is not to be considered in assessing the wrongness of the perpetrator’s act.

Proponents of hate/bias crime legislation respond to this worry by urging that hate/bias crimes are different because they are aimed at groups and not just at individuals. Yet in worrying about wrongdoing, as opposed to culpability, this response is surely beside the point. If collateral harms to those who are not themselves crime victims are relevant to gauging the degree of wrong done by an offender, it will not be because of the offender’s aims. As long as offenders have some culpable mens rea with respect to such collateral harms, that should be sufficient for us to blame them for such harms, if such harms are otherwise relevant.

Our view is that such collateral harms are generally relevant to assessing the wrongfulness of an offender’s actions, and may be taken into account either in the definition of the actus reus of more serious grades of offenses, or in the appraisal of victim impact statements at sentencing. When we set severe punishments for assassinations of public officials, or assaults upon police officers, we take into account the collateral harms (to persons other than the direct

---

38 Id. at 42, 63.
victims of those crimes) caused by such acts. When we allow grandmothers of murder victims to submit victim impact statements at the sentencing of offenders\(^\text{39}\) we are inviting jury consideration of collateral harms in assessing the degree of wrong done by those convicted of murder. Sentence enhancements for hate/bias crimes thus might be based on the collateral harms that those crimes cause.

There are more serious problems for the collateral harms argument for sentence enhancements, however. Such problems parallel the problems raised for the previous argument concerning the uniquely severe psychic trauma caused to the principal victims of hate/bias crimes. First, there is the nagging question of empirical fact: is it really true that members of groups suffer more vicarious psychic injury in response to hate/bias motivated crimes than in response to otherwise-motivated crimes? The question, again, is not whether such vicarious suffering exists at all; it is, rather, the comparative question of whether it exists to a significantly greater degree within groups targeted by hate/bias crimes than it does within groups related in various ways to crime victims generally. As we saw before, the social science literature on this topic—even for principal victims of bias crimes—is not very reassuring. And if we must make the judgment from the armchair,\(^\text{40}\) our bets are with John Donne, who famously intoned that when the bell tolls for any of us it tolls for all of us, be one white, black, Catholic, Jewish, gay, straight, disabled, male, female, etc.

\(^{39}\) In *Payne v. Tennessee*, 501 U.S. 808 (1993), the grandmother described the trauma of her grandson in reliving the witnessing of the deaths of his mother and sister at the hands of the defendant.

\(^{40}\) For some speculations about this, see Jacobs and Potter, *supra* note 4, at 87.
Second, there are the same proxy problems for vicarious psychic trauma arguments as for the direct versions of these arguments. Again, the hateful or biased motivations of an offender are a poor proxy for the vicarious psychic trauma of members of the victim’s community. The group’s perception of hate/bias would be a better proxy. In addition, even perceived hate/bias is a proxy that is both over- and under-inclusive. It is also quite unnecessary. If severe collateral trauma is caused by an offender’s crime (whatever its motivation), that harm can be taken into account via victim impact statements at sentencing. There is no need to enhance the sentences of all offenders who commit crimes because of hatred or bias simply because some of them cause greater, independently-provable harms to their victims (including those who are victims vicariously). Those who culpably cause such harms (whatever their motivation for so doing) ought to pay for them with greater sentences; those who do not cause such harms ought not to be penalized because others do.

The case for the enhanced sentencing of hate/bias-motivated crimes based on secondary psychic trauma is thus even shakier than was the case based on the psychic trauma of principal victims of hate/bias crimes. Neither the factual premises nor the normative inferences implicit in these arguments can be sustained.

4. Social Harms

Lastly, proponents of hate/bias crime legislation urge that all of society is harmed by such crimes, not just the immediate victims and the groups to which they belong. There are three sorts of harms with which advocates of this “social harm thesis” appear to be concerned. First,
they claim that hate/bias crimes have a chain reaction effect, so that members of the group who feels attacked are likely themselves to attack members of the group with which the hate/bias crime offender is associated. Let us call this the “blood feud wrong,” since it invites the picture of a violent circle of retaliatory violence\(^{41}\) by Appalachian families long at odds with one another. Secondly, social harm theorists claim that hate/bias crimes cause a general breakdown in citizens’ sense of order and security. As Weisburd and Levin claim, “criminologists have found that among individuals who face little direct threat of criminal victimization there is a fear and anxiety that is related to a breakdown of community order and civility in their surroundings.”\(^{42}\) Thirdly, hate/bias crimes are said to increase the polarization of a society. Groups feel more isolated and less akin to one another, and their commitment to the egalitarian ideal is thereby weakened.\(^{43}\)

This grabbag of arguments should be seen for what it is, a makeweight. For the harms claimed here are so abstract, the causal connection of hate/bias crimes to such harms so attenuated, that no one could seriously be motivated to support hate/bias crime legislation on these bases. Let us briefly consider each of them in turn.

The first argument is reminiscent of the old “prevention-of-vigilante-justice argument” for punishment generally.\(^{44}\) That old argument is that we should punish those who are thought to deserve punishment because if we do not do so, vengeful citizens will take the law into their own hands. Here, we should enhance the punishment of hateful or biased offenders because, if we

\(^{41}\) Weisburd and Levin, *On the Basis of Sex*, supra note 3, at 26; Lawrence, Punishing Hate, *supra* note 3, at 42.

\(^{42}\) Weisburd and Levin, *supra* note 3, at 26-27.

\(^{43}\) Lawrence, Punishing Hate, *supra* note 3, at 43.

did not do so, the victimized group would retaliate and become a victimizing group. Such claims are both factually implausible and normatively undesirable. After all, it is hard to lay at the door of some offender a harm caused to him (or to other members of his group), and use that harm as a reason to punish him more. The people who need punishment are those who retaliate, if any do. In short, the solution to the problem of vigilanteeism is to punish vigilantees.\textsuperscript{45}

The second argument is reminiscent of Lord Devlin’s argument for why we should criminalize homosexual behavior: if we do not do so, society’s sense of decency and order will give way and, like treason, homosexuality will lead to society’s collapse.\textsuperscript{46} Similarly, inadequately punished hate/bias crimes will so weaken our sense of order, civility, community, etc., that we will face social breakdown. No one believed such hysterical claims about why homosexuals should be punished, and no one should believe that they furnish a legitimate basis for increasing the punishments of those who attack homosexuals. All crime cuts into our sense of order, our sense that society harmoniously works, our faith that the social contract is being kept, etc. There is nothing that can be said about the corrosive effects of hate/bias crimes on the social order that can not be said about other crimes.

The third argument could furnish a plausible justification of hate/bias crime legislation, depending on a society’s reaction to hate/bias crimes. It is possible for our sense of community to be diminished by such crimes; but it is at least equally possible for our sense of kinship to be kindled by such crimes. A common saying in legal philosophy is that Hitler converted more people to natural law than Thomas Aquinas ever did. Similarly, the New Bedford rapists may

\textsuperscript{45} See Moore, Moore, Law and Psychiatry, \textit{supra} note 5, at 235-236.

\textsuperscript{46} Patrick Devlin, The Enforcement of Morality (Oxford, 1965).
have increased the solidarity across the sexes, as evidenced by the vigils held across the country by people of both sexes after the event. Given the selection of the most brutal of hate/bias crimes for extensive reportage—such as those in Texas\textsuperscript{47} and Wyoming\textsuperscript{48}—surely most people’s aroused revulsion operates to increase, not decrease, the identification with members of different races, ethnicities, ages, sexual orientations, etc.

It is no accident that each of these three “social harm arguments” harken back to general arguments made to support having a criminal justice system at all. That is because these very abstract concerns are not uniquely raised by the prospect of under-punishing hate/bias crimes. Revenge by citizens, a diminished sense of order and security, a diminished sense of community and equality, would be the general costs of not having a system of criminal law and of punishment. To invoke them to justify increasing punishment for a particular type of crime is to commit as great a category mistake as that made by Devlin in invoking like arguments to justify criminalizing specific types of unpopular behavior of minority groups.

C. Motives of Hate as Constituting Greater Deontological Wrongs

As we mentioned in the earlier discussion on the concept of wrongdoing, the conception of wrongdoing in terms of causing those bad states of affairs we call “harms” is only one possibility. The other possibility is to conceive of wrongdoing as an act that instantiates a categorical norm of obligation. On a non-consequentialist view of morality, wrongful action

\textsuperscript{47} Cite.

\textsuperscript{48} Cite.
does not consist of doing something productive of bad states of affairs (where the badness of the state of affairs is what makes the action wrongful). Rather, the gravamen of a moral wrong (on this alternative view of morality) consists simply in either doing an act prohibited by morality or not doing an act required by morality.

On a non-consequentialist, or “deontological” view of morality, an action can be right or wrong intrinsically, that is, without regard to the badness of the state of affairs such as an action may cause. It is this independence of wrongfulness from harms or any other bad states of affairs that opens up the possibility of there being greater wrongdoing despite the absence of any greater harm caused by an action.

On the deontological conception of wrongdoing it might seem that all wrongs are equally wrong and thus that there cannot be an idea of greater or lesser wrongdoing. After all, on the deontological view, wrongdoing consists of violating a categorical norm of obligation – and those norms (for a deontologist) categorically forbid the doing of various actions. If we are categorically forbidden either to kill or to steal, how can there be any greater wrong in killing as opposed to stealing? Surely, one might think, we shouldn’t do either, no matter what the consequences.

There is something of a puzzle here for deontologists, leading to some interesting conundrums explored by Leo Katz.49 For our purposes, however, perhaps the following will suffice. Greater wrongdoing consists in the violation of a more stringent moral norm. Although all such norms are categorical, some are more stringent than others. E.g.,: the norm against killing is more stringent than the norm against stealing. How the idea of stringency is to be

49 Leo Katz,
cashed out differs amongst deontologists: some think that categorical norms can conflict in the actions they forbid or require, and that stringency is a function of winning in such cases of moral conflict; others think that moral norms come with exceptions to them, and that stringency is a function of the numbers and scope of the exceptions existing for a given norm; others, “threshold deontologiss,” think that sufficiently good consequences can justify the violation of otherwise categorical norms, and that the stringency of a norm is measured by the heights of the threshold where good consequences can do such justifying work; still others urge that stringency is measured by duties to prevent and/or repair.

We will here simply assume that the idea of greater and lesser stringency can be given one or more of these senses. With such assumption the deontological notion of greater wrongdoing is simply stated: such wrongdoing consists in the violation of a more stringent norm of obligation. With this concept of greater wrongdoing in hand, let us turn to possible applications of this concept to crimes motivated by hate.

Because some crimes like assault violate the rights of victims of such assaults, it is plausible to suppose that an assault, whatever its motivation, violates a categorical obligation and is, in that deontological sense, wrong. A proponent of hate crime legislation must then think that the motivation of such an assault by hatred violates a more stringent obligation of morality and

---

50 Kant famously thought that any such conflict of obligations was inconceivable. A deontologist who is also a Kantian in this regard will not cash out the idea of stringency in this way.

51 See Moore, Placing Blame, supra note 5, at 717-724.

is, in that, deontological sense, a greater wrong. The challenge for such proponent is to specify
the content of such more stringent moral norm that hate-motivated crimes violate.

The possibilities here are large, depending on one’s view of the content of morality’s
norms. Let us strive for a semblance of order by grouping the possibilities into two general
positions. The first we shall call the Kantian position. On this view, the categorical norms of
morality are aimed at reasons as much as actions. Our moral duties are not just not to do certain
actions; they also are not to do such actions for certain reasons. Because of this two-level
structure of obligation, if we do some act A violative of our duties that is wrong; but it is more
wrong to do A for reason R, if R is a prohibited reason.

The application of this (broadly) Kantian view to hate crime legislation should be
apparent. Although it is wrong to assault another, it is a greater wrong to assault another for bad
reasons such as racial hatred. So long as hatred and bias are amongst morality’s prohibited
reasons, the conclusion of greater wrongdoing being involved with hate motivated crimes falls
out naturally from the Kantian view.

Or so it seems. In fact when Kant’s view of morality makes reasons the object of
morality’s prohibitions, Kant was talking about ends (or goals). His was not the view that
morality’s norms of categorical obligation prohibit either the having of, or the acting on, certain
emotions like hatred or certain predispositions like bias. These aspects of one’s character may be

---

53 We are unconcerned with the historical question of whether Kant actually held this position. For a contemporary philosopher who holds the view in the text, deriving it in part from his reading of Kant, see Joseph Raz, Practical Reason and Norms (Oxford, 1975).

54 Kant
the concern of an Aristotelian view of morality, but they were not for Kant the locus of our obligations.

As we discuss in the subsequent sections on culpability, Kant was onto something in making the distinction between reasons as goals and reasons as emotions/dispositions. We can choose our goals; we have limited capacity to choose our characters, including the components of character such as what we feel and what we are disposed to believe. Thus, on the supposition that obligation exists only vis-à-vis what we can do something about, it is implausible to think we are obligated to feel differently – even if Kant is correct that we are obligated to choose different ends for our actions.

Unfortunately for proponents of hate crime legislation, Kant is not correct even on his own turf: we do not have categorical obligations with respect to our motives. Our categorical obligation is not to kill (with some exceptions); we have no obligation not to kill for revenge, for example, in any sense save as an instance of the obligation not to kill, full stop, for any reason or for no reason.55

That the categorical norms of morality are directed at actions, and not those actions’ reasons, is evidenced by the structure of all criminal codes. Such codes prohibit the actions of killing, raping, stealing, etc. With rare exceptions such codes do not prohibit “acting out of greed,” “acting from base motive,” etc. While Kant took this feature of the criminal law to evidence no more than the administrative needs of the law, it goes much deeper than that. These codes evidence the fact that the morality of obligation – the kind of morality with which criminal law deals – is about actions and not motives.

55 We each have argued this at some length. See Michael Moore, Educating Oneself in Public 168-169 (Oxford, 2000); Heidi Hurd, What in the World is Wrong?, supra note 16.
The conclusion that our obligations are about actions and not reasons, is buttressed by the fact that we cannot choose not to act on certain reasons as we can choose not to act. We can (as we noted before) choose our goals (or ends); this distinguishes such goals from emotions, which we cannot will into or out of existence. Yet we cannot will that an end motivate, or not motivate, a certain action. In that sense we cannot choose our reasons for an act, even while conceding that we can choose whether to do the act. On the supposition that morality cannot plausibly be thought to obligate us to do what we have no power to do, it follows that moral obligations concern our actions and not their motives.

If the moral geography of obligation is as these two pieces of evidence suggests that it is, then there is no greater wrong done in hate-motivated crimes than in ordinary, other-motivated crimes. There may be greater culpability – see Part III below – but there is no greater wrongdoing to be found here.

The second, non-Kantian position (about greater non-harm-based wrongdoing being involved in hate crimes) lies in what Frances Kamm has called the “organic harm thesis.”\textsuperscript{56} In Kamm’s language, such organic wrongs “are ‘organic’ in the sense that they are greater than the sums of their parts; this in turn means that they are multiplicative, rather than additive functions of their parts. That is, we do not treat a burglary caused by hate of a race as if it involved simply the presence of hate plus the occurrence of a burglary. Rather, the fact that the hate caused the burglary is supposed to make the burglary worse.”\textsuperscript{57}

\textsuperscript{57} \textit{Id.} at 630.
The arithmetical metaphor actually obscures Kamm’s real point about organic wrongs. This is the point that “factors placed together may not merely add up, but also interact with one another.” Her example is the criminal who shoots the victim with the victim’s own gun. This is not, Kamm thinks, just the wrongs of assault (of a person) and theft (of the gun). It is also a distinct, and greater, wrong of its own: “it is worse to be harmed by the use of what is yours, and what should be used for your benefit, than it is to be harmed in some other way.”

What Kamm’s example reveals is how the organic harm thesis does not concern the hatred or bias of the hate-crime perpetrator. These are not what interacts with the wrong that is the base crime to produce a greater wrong. Rather, it is some objective circumstance in the presence of which the doing of the base crime becomes a greater wrong. In her example, that circumstance is the fact that the gun used on the victim was the victim’s own gun. This is not part of the mental state of the perpetrator but is rather an objective feature of the action done by the perpetrator.

The circumstances mentioned in the literature favoring hate crime legislation run the gamut, including: the greater vulnerability of typical hate crime victims; the fact that the characteristics by virtue of which hate crime victims are selected for attack are immutable characteristics; the fact that our history includes many similar acts motivated by racism, homophobia, sexism, etc., so that any new hate crime is a form of worse-because-repetitive-wrongdoing; the fact that other potential victims just like hate crime victims (save for their races, ethnic origin, sexual preference, and the like) are not subject to these sorts of attacks; the fact

58 Id. at 631.
59 Id. at 631.
that many persons view hate crimes as expressing a message of contempt for the groups from which hate crime victims are selected.

We shall want to say a bit about each of these circumstances separately, but the first thing to say about all of them is that they are objective features of the situation in which some hate crimes take place; they are thus not themselves bad motives of the perpetrators of hate crimes. So if these factors do increase the wrongdoing done, enhancing punishment based on bad motives of hate or bias is an indirect means of punishing for this greater wrongdoing. Bad motive is again being used as a proxy for these features, and the question recurs as to how such a proxying function is to be justified. If vulnerability of the victim makes for greater wrongdoing, for example, why not generally enhance the punishments of criminals who pick on vulnerable victims? Why single out hate-motivated criminals, a proxy we know is wildly under-and-over-inclusive for vulnerability victims? Surely something like the current federal sentencing guidelines more accurately take this factor into account when they increase punishment for any crime where the “victim of the offense was unusually vulnerable due to age, physical or mental condition, or . . . otherwise particularly susceptible to the criminal conduct . . .”\(^{60}\)

All non-Kantian forms of the thesis of greater non-harmed-based wrongdoing are thus met at the outset with our (now repetitive) proxy objection: if one really thought that these features make for greater wrongdoing, why not punish such features directly rather than by the proxy of bad motive? While we find the proxy objection sufficient here as before, we shall nonetheless turn to the features being proxied. For there are questions about their ability to enhance wrongdoing even if they were used directly as punishment enhancers.

Take heightened vulnerability of victims first. There are two ways one might suppose such factors to justify increased punishment. One is by the greater harms more vulnerable people suffer when they are the victims of crimes: “vulnerability” can be cashed out as susceptibility to physical or psychic injury, in which case crimes to more vulnerable victims are crimes causing more physical and psychic injury. But this returns us to harm-based wrongdoing and on our previous objections to that version of the thesis.

The other route to justifying increased punishment for crimes done to vulnerable victims, is through the cowardice such crimes evidence. Picking on the particularly gullible, stupid, ignorant, old, handicapped, very small, weak, youthful, insane, etc. victim may seem cowardly because there is no fighting chance to get the better of his attacker possessed by the victim. To use an old phrase, there is no honor in such crimes, even for criminals with their peculiar sense of honor.

We have some sympathy for this view of the relevance of vulnerability to sentence enhancement. But notice now how broad must be the notion of vulnerability to accurately capture this insight: all of us, Arnold Schwartzenegger included, can be particularly vulnerable to crime. If we turn our back at the wrong moment, are found unconscious, are tied up, etc., we are the easy victims no “honorable” criminal should touch. Any of us can become such easy marks that there is no honor in making us the victims of crimes. Now the proxy objection becomes decisive: there is no justification whatever for using bad motives of hate and bias as a proxy for this very general potential for vulnerability. The preferable statute enhances punishment directly for vulnerability. Moreover, where there already are punishment enhancements for vulnerability of victims, as in the federal sentencing guidelines, it would be double counting to enhance
punishment still further because of this same factor. Proxying the factor of vulnerability disguises the factor but it doesn’t change it to prevent double counting.

Lawrence Crocker has advanced the thesis that it is the immutability of the characteristics by virtue of which hate crime victims are singled out for crime that is “part of the answer” as to why hate crimes are worse (in a non-harmed-based sense). One could view the immutability of such “victimogenic” characteristics as one kind of vulnerability: persons possessing such characteristics can do nothing about them because they are immutable characteristics. Yet this is a non-starter of an argument. Many characteristics besides race, gender, ethnic origin, are immutable, and many of these render us far more vulnerable to crime: size, strength, age, mental defects, etc. Moreover, the immutability of a characteristic is a tiny part of what makes the possessor of it vulnerable to crime. Even if we could have changed some characteristic making us vulnerable, if we haven’t made such a change at the time of the crime we are in fact more vulnerable (and a criminal more cowardly for picking on us).

Crocker in fact does not raise immutability as a route to vulnerability. Rather, he takes immutability to negative any inference of victim provocation or fault: “often assaults are provoked by behavior with which an ordinary person could take some exception. Assaults provoked by immutable characteristics seems unjust in comparison.” Yet surely this too is an argument with no chance of surviving even modest scrutiny. It boggles the mind to think that if one of us is assaulted because he is rich, or because she wears a low cut dress, or because he speaks his mind on some issue of public importance, or because he performs abortions, etc., etc.,


62 Id.
that because these characteristics were mutable by the victim in any sense makes his/her victimhood partially their fault. We are entitled by right not to change these admittedly changeable characteristics. It is thus no less unjust for us to be assaulted because of such characteristics than for a hate crime victim to be assaulted because of some immutable characteristic. Characteristics we cannot change, and characteristics we should not have to change, are equally non-provoking characteristics for which crime victims are not at all at fault for possessing.

Crocker’s main suggestion for why hate crimes are worse is the third of those earlier enumerated: “hate crimes are worse, independent of consequences, because of the history of racism, and the other enumerated hatreds.” Crocker’s suggestion is built on the idea that a given crime can be made worse by being preceded by other crimes done earlier by that criminal. Crocker appeals to the widespread support for enhanced punishment for recidivists on the basis that doing the same wrong a third, fourth, or fifth time is worse than doing it for the first time.

Enhanced punishment even for true recidivists is problematic from a retributivist point of view. Yet these problems are greatly exacerbated when one enhances punishment for those who repeat the crime of others. There are two sorts of desert-based justifications for enhanced punishment of true recidivists. One appeals to the worse character of the true recidivist: unlike the one-time offender, the recidivist both has, and can be more easily proven to have, a bad character. His last crime cannot be seen as an aberrational act, out of character for him; rather, such act expresses his true nature.

63 Id.

64 For exploration of how one might justify enhanced punishments for true recidivists, see George Fletcher, Rethinking Criminal Law 460-466 (Little, Brown, 1978).
When the past acts are not those of a given defendant, however, this inference to bad character is blocked. If anything, the existence of a history of past acts of the kind done by some present defendant makes it somewhat more likely that his action does not reflect his true character; perhaps he was merely succumbing to stereotypes learned from knowing that past history. Parroting such history does not betoken bad character so much as it betokens weak, only partly formed, character.

Crocker, seeing the problem, concedes that “bad acts committed by others before the birth of the criminal actor surely do not tack on to this person’s own acts.” Yet in a sense Crocker thinks they do: “To act from a racist motive is in part to ratify that (racist) history, to make it one’s own through a concrete act.” Yet in what sense do the acts of earlier racists become “one’s own” through imitation? How does one “ratify” an earlier criminal’s action so as to place that action on one’s own moral ledger, in any sense of that phrase? Such claims are reminiscent of Jean Paul Sartre’s old notion that we should accept the responsibility for all of the world’s evils. It is all very well to urge an acceptance of the task-responsibility to do something about such evils, including racist evils; but it boggles the mind to think that we can make ourselves responsible for those evils by any act of contract-like ratification, no matter how explicit. We can make ourselves responsible for evils we did not cause no more than we can, after the fact, make home runs out of pop flys. To think otherwise is to suffer from some delusion of grandeur and omnipotence typical of some forms of paranoia.

---

65 Crocker, supra note 61, at 493.
66 Id.
67 Jean Paul Sartre, Existentialism is a Humanism, in Existentialism versus Marxism (Nowack ed., 1965).
The second route to justifying enhanced punishments for true recidivists does not depend on some insight into bad character offered up by repeated criminal acts. Rather, this route focuses on the culpability with which the most recent act of wrongdoing is done. It is said that a repeat criminal has every chance to learn the wrongness of his acts. His own guilt responses, his empathy for his victims, the negative reactions of others to his crimes, should show him the error of his ways. If he persists in crime despite these enhanced opportunities to know that what he is doing is wrong, then he is more blameworthy than is one lacking such enhanced opportunities.

This argument holds more promise of transferring to those who imitate the earlier crimes of others. For we can get some of such knowledge by our attention to the earlier wrongs done by others, if we know about them. And perhaps it is true, as Crocker argues, that “it is hard to believe that anyone growing up in modern society could fail to be aware of the evil of . . .” doing violence to the person of another for no good reason. Everyone surely has as much opportunity to acquire this moral knowledge as to acquire the moral knowledge about racism of which Crocker makes so much. The ordinary assaulter, like the racist assaulter, has enough precedents to liken him to true recidivist assailters in this respect.

68 Crocker, supra note 61, at 493.
While we find these arguments dispositive of Crocker’s recidivist analogy, we also do not share Crocker’s belief in the cogency of these arguments even as they apply to true recidivists. The first argument supposes that it is character that is the true desert basis for punishment, bad acts being only the law’s convenient marker for bad character. The second argument presupposes that knowledge of wrongness is relevant to the culpability of a wrongdoer. In earlier work we have sought to dispute both of these presuppositions.69 If our earlier conclusions are correct, then even true recidivists cannot justly have their punishments enhanced for these reasons. *A fortiori*, the same goes for racist ratifiers of the earlier wrongs of others.

The fourth aspect of hate crimes that allegedly makes for greater wrongdoing is the discriminatory nature of hate crimes. Hate crimes victims are not just injured by a hate-motivated assault; they are said to suffer the additional wrong of being discriminated against.70

This claim can be dealt with briefly because it is such a non-starter. It is true that no one deserves to be a crime victim because of their race, ethnic origin, etc. That someone was picked out when others, equally situated, were not, is a kind of unjust discrimination; such crimes violate our norm of equality as well as our substantive norms against unjustified violence. Yet this is true of all crime victims. No one deserves to be singled out for violent crime for any reason; when a victim is so singled out s/he always can truthfully say that s/he is in no morally relevant respect different than others not so targeted and concludes therefore that her right to equality was violated as well as her substantive right to bodily integrity. Hate crime victims thus

69 On the use of character as the desert-base for responsibility, see Moore, Placing Blame, *supra* note 5, ch. 13. On the requirement of moral knowledge for responsibility, see *id.* at 412-413.

70 Anthony Dillof explores this possibility in depth, concluding as do we that it cannot constitute the alleged greater wrongdoing incident to hate crimes. See Dillof, *supra* note 4, at 1037-1049.
can tender no special violation of their rights to equality not equally claimable by every other crime victim.

The fifth and last claim of non-harm based wrongdoing is one focusing on the hateful message said to be expressed by hate crimes. The argument here is that a separate wrong – a “mimetic wrong” – is done by the explicit saying of such a hateful message. Because this version of the “greater wrongdoing” thesis involves an entire approach to law that has come to be known as “expressivism,” we separate this claim out for more extensive treatment in the next section.

II. THE EXPRESSIVIST ANALYSIS

A. The Two Forms of Expressivist Theories

Probably the most popular argument offered in support of hate/bias crime legislation has been put in (what has come to be called) expressivist terms. As summarized by Dan Kahan, a leading proponent of the expressivist theory of punishment: “From an expressive point of view, hate crime legislation can be used to criticize the devaluation of gays... [and other groups].” 71 The expressivist argument is two-fold. First, we are to view criminal activity as expressing a message–that of devaluation of the victim and the class of which the victim is a member. As Kahan writes: “On expressive grounds, serious crimes strike us as such – that is, as crimes and as serious – not just because they impair another’s interests but because they convey that the

wrongdoer doesn’t respect the true value of things.”72 Second, we are to view criminal punishment as expressing another message, namely, a message disapproving and denouncing the message sent by the criminal. In Kahan’s words: “To express condemnation, then, society must respond with a form of punishment that unequivocally evinces the community’s repudiation of the wrongdoer’s valuations.”73 On the expressivist view both crime and punishment are seen as communicative exercises. The criminal opens the conversation by saying, via his crime, “I don’t value you and those like you as persons,” and society answers back via its punishment, “you are wrong not to value such persons.”

There are two ways to construe the expressivist analysis of hate/bias crime legislation.74 On one interpretation, the expressivist thesis is at home with conventional theories of punishment, such as retributivism, deterrence-based utilitarianism, or a mixed theory. On this construal, the enhanced punishment of hate/bias crimes is justified on the ground that criminal acts which send hateful messages are more wrongful than are criminal acts that do not send such messages. Not only do hate/bias crimes invade victims’ interests in bodily security, the protection of property, and mental tranquility, but they further wrong victims (and their communities) by communicating false and hurtful messages about such victims, and such further wrongdoing merits greater punishment under standard punishment theories.

The second way to construe the expressivist justification of hate/bias crime legislation is more radical. This construal abandons the traditional retributivist and utilitarian theories of

73 Id.
74 Kahan has sympathies with both construals of expressivism. See Kahan, What Do Alternative Sanctions Mean?, supra note 71, at 601.
punishment in favor of an altogether new theory of criminal justice. On this new theory, punishment is justified if and only if it sends an appropriate message of denunciation to a defendant and those who share the defendant’s corrosive views. The point of punishment, on this view, must be to contradict the disrespectful messages conveyed by criminals. Those who defend this second interpretation conceive of legally-conveyed denunciation as such a good thing that it and it alone can justify punishment. Frederick Lawrence quotes Lord Denning’s famous statement of this view in 1954: “The ultimate justification for any punishment is . . . that it is the emphatic denunciation by the community of a crime.”\(^75\)

Both the “moderate” and the “radical” construals of the expressivist justification of hate/bias crime legislation threaten to collapse back into more traditional claims. The danger for the first, or “moderate”, interpretation of the expressivist thesis is the likelihood that what makes the criminal’s expression of his hateful message wrong is just that it will lead to the kinds of harms that we have already canvassed in more traditional terms. If, for example, the wrongness of hateful expressions resides in the fact that such expressions induce more assaults, or more violent assaults, or greater psychic injuries to principal victims or to those who share the characteristics of principal victims, or greater social unrest generally, then the so-called expressivist wrongs perpetrated by those who commit hate/bias crimes are no different than the kinds of harm-based wrongs considered in Part I. These claimed invasions of physical, psychic, and social interests were both better described in the non-expressivist terms of the arguments of Part I, and in any event suffer the same problems as did those arguments. The pressure is thus on

\(^75\) Lord Denning’s statement was made as author of the Royal Commission on Capital Punishment and is much discussed in H.L.A. Hart, Punishment and Responsibility (Oxford, 1968). Denning is quoted approvingly in Lawrence, Punishing Hate, \textit{supra} note 5, at 163,
those who advance this first version of the expressivist thesis to come up with some account of
the wrongness of hate- and bias-induced crimes that does not reduce to the accounts we have
already found wanting.

The danger that confronts the second “radical” interpretation of the expressivist thesis is
that denunciation of crime will turn out to be good only inasmuch as such denunciation serves
either utilitarian or retributivist goals. That is, if denouncing crime is good because it educates
citizens not to engage in such activities and reaffirms the social contract, and if these things are
good because they reduce crime, then a denunciatory theory of punishment is but a disguised
form of a utilitarian theory of punishment. Alternatively, if denouncing crime is good because it
shames criminals, and if this is good because shaming them makes them suffer in a manner that
is deserved, then a denunciatory theory is but an application of a general retributivist theory of
punishment. The pressure is thus on the “radical expressivist” to advance an account that makes
plausible the intrinsic goodness of denouncing criminals—a goodness that could justify their
punishment without being in the service of the traditional penal goals of either crime prevention
or the giving of just deserts.

In what follows we shall take up, in turn, each construal of the expressivist justification
of hate crime legislation. As we do so, we shall focus on the described need for expressivists to
state their theories in ways that make them truly different from the traditional theories of
wrongdoing and of punishment that we have already applied to the analysis of hate/bias crime
legislation.

B. The Expression of Hate and Prejudice as a Unique Form of Wrongdoing
The challenge to the moderate expressivist who seeks to work within the traditional theories of punishment (retributivist, utilitarian, or mixed) is to come up with some account of why hate- and bias-motivated crime makes for greater wrongdoing in a way that does not collapse the argument into the more traditional wrongdoing thesis that we already rejected in Part I. Whether the expressivist can meet this challenge depends on the sense to be given to the idea that crimes express something possessed of bad social meaning.

The most obvious ways of construing how crimes can express a social meaning of hate or prejudice are not open to the expressivist if he is to meet the challenge mentioned above. For the most obvious ways of construing how hate/bias crimes possess “social meanings” that “convey” or “express” devaluation of others all stem from a basic picture of communication. On a standard view, successful communication involves three elements. First, one who communicates means something by his action. J.L. Austin referred to what an actor means as the actor’s “locutionary act;” linguists usually call it the “speaker’s meaning;” and one of us has elsewhere dubbed it the actor’s “semantic intentions.” Distinct from the semantic intentions of a speaker are both his further intentions in doing the act – what the speaker means to achieve by his action – and, in Austin’s language, his “illocutionary” intentions vis-à-vis the kind of action he is doing, such as

---

76 J.L. Austin, How to do Things with Words (1965).
asserting, questioning, commanding, etc.\textsuperscript{79} Although distinct, an actor’s further and illocutionary intentions are usually helpful in assessing the semantic intentions of a speaker.

Second, although one could mean, “It is snowing in Tibet.” by uttering “Gleeg, gleeg, gleeg!,”\textsuperscript{80} a rational speaker who wishes to communicate the thought will choose more conventional means of expressing it. This is because communication is a kind of co-ordination game between a speaker and an audience:\textsuperscript{81} the speaker wishes her audience to understand her thought, and the audience wishes to understand that thought as well. What is needed is some salient convention onto which both the speaker and the audience can fasten. In linguistics such conventions are often (if controversially) divided between semantic conventions (giving the meaning of words and sentences) and pragmatic conventions (giving the norms of appropriate utterance). Whatever the details, the general point is that communication succeeds through the use of social conventions shared by speaker and audience.

Finally, the last element of successful communication is the belief of the audience. Beliefs, like intentions, are propositional attitudes, and they thus join semantic intentions and linguistic conventions as the bearers of meaning. When the beliefs of the audience match the intentions of the speaker – so as to achieve what Austin called audience “uptake”\textsuperscript{82} – then communication has succeeded.

\textsuperscript{79} Id. at 339. See also Proceedings of the Law and Linguistics Conference, 73 Wash. U. Law Q. 800, 886 (1995)

\textsuperscript{80} Paul Ziff’s example in arguing against speaker-oriented theories of semantics. See Ziff, On H.P. Grice’s Account of Meaning, 28 Analysis 1, 5 (1967).

\textsuperscript{81} David Lewis, Convention (1970).

\textsuperscript{82} J.L. Austin, supra note 76, at ______.
Because expressivists view criminal actions as conveying a message – as expressing something possessed of social meaning – one’s first temptation is to take them at their word and apply the complete model of communication to criminal actions. Yet this plainly will not do. First, it is obviously false that criminal actions communicate a message of hatred, disrespect, or devaluation in the tripartite sense discussed above. Consider this old example: “A raised his voice because he was angry.”

It is possible that what A did constituted an act of communication: he intended to express the fact that he was angry because he wanted people to know this fact; and he raised his voice in order to get them to arrive at this belief. (He may even have intended to induce knowledge on their part that he wants them to know that he is angry.) But such a scenario is surely a rarity. Usually when someone raises his voice because he is angry, there is no communicative act going on at all. The raising of the voice is the effect of the speaker’s anger, but it is not the object of any intention on his part to communicate that anger by raising his voice.

Similarly, criminal actions directed against certain victims because of their race, sexual orientation, gender, etc., could sometimes be communicative acts because the criminal seeks by them to send a calculated message of hatred or prejudice; much more often, however, hate-motivated crimes have no such communicative intention behind them, even though they are hate crimes, viz, crimes done because of hate. To avoid obvious falsity, thus, the expressivist must construe the “conveying of hateful messages by criminals” in some way other than by use of a full blown theory of communicative actions.

---

83 From Moore, Law and Psychiatry, supra note 5, at 15-18, 346-47.
The natural move is to pry apart the three elements of a full-fledged theory of communication and to use only one such element as the locus of the “social meaning” that criminal actions “express.” Consider the first such element—the intentions of the actor. As we have just seen, those guilty of hate crimes typically have no semantic intentions – they do not mean something by their acts – nor do they mean to communicate that something, in the sense of induce belief in their victims or others. So if the “social meaning” such crimes express is to be found in the subjective mental states of criminals, it cannot be in terms of semantic and communicative intentions they do not typically possess.

Still, hate- and bias-motivated criminals clearly possess non-communicative mental states of belief and emotion. These are the mental states such criminals must possess if they do their crimes because of the race, ethnicity, gender, etc., of their victims. Perhaps the social meaning such crimes express is to be found in offenders’ cognitive and motivational states. This hope should be abandoned, however. For if one thinks that the emotions or stereotyped beliefs constitutive of hate/bias crimes make offenders more blameworthy, that thought can be fully accommodated in non-expressivist terms. One might think such beliefs or emotions make a criminal actor more blameworthy because they make him more culpable (an argument that we shall examine in the next Part): alternatively, one might think that their greater blameworthiness is due to their contribution to greater wrongdoing. But even when it is the latter possibility on which one is focusing, such greater wrongdoing would be of the organic harm sort examined in the last section. That is, the expressivist claim would be that the subjective beliefs and emotions of the hate-motivated criminal makes his act a greater wrong because his obligations are directed against certain reasons for actions as well as against certain actions themselves. This
expressivist claim is no different (and no better) than the non-expressivist claim we examined before.

The next move open to the expressivist is to repair to the third element of a full-fledged theory of communication, the element of audience belief. There is nothing nonsensical about attributing a meaning to an action solely on the basis of how an audience understands that action. Audience-centered interpretive schemes are both well established and make good sense for defamatory utterances, for example, and (to a lesser extent) for contracts, statutes, and constitutional provisions, as well. So perhaps the social meaning of hate crimes is to be found in the beliefs induced by such crimes in their principal victims or in others.

A minor problem for this audience-centered construal of the expressivist’s thesis is that not all hate/bias crimes actually cause the beliefs in victims that are needed to invest the criminal act with the “social meaning” assigned to such crimes by expressivists. Some victims of hate crimes undoubtedly do not perceive the hatred behind their assaults – e.g., being hit in the back of the head by an unknown but hateful assailant – and neither may anyone else. So some hate crimes will not express anything capable of having a social meaning, if this is what “social meaning” means.

A much more important problem is the redundancy problem. Even when victims of hate crimes or others do feel the sting of the hate or prejudice– e.g., they well know of the perpetrator’s hatred of certain groups, including some to which they belong – we need to ask why this understanding by victims or others is normatively important. More specifically, what is it

---

84 W. Page Keeton et. al., supra note 35, at 808-809.
85 Proceedings, supra note 79, at 887-888.
about this consequence of hate/bias crimes that makes such crimes deserving of greater punishment? Surely the only plausible answers are those we examined in Part I in the non-expressivist terminology of interest invasions and harms. Such hatred and bias is a proxy for greater physical harms to the victim, or for greater psychic harms to the victim, or for greater psychic harm to the group that is the object of the criminal’s hatred, or for the greater social harms of unrest, class warfare, and vigilante justice. These are simply the four sorts of harms allegedly brought about by hate- and bias-motivated crimes, as distinct from otherwise-motivated crimes. If a crime’s “social meaning” is thus equated with the beliefs of its audience, the expressivist adds nothing to the non-expressivist justification for hate/bias crime legislation advanced by harm-based wrongdoing theorists.

The third possibility for the expressivist is to take the conventions that make communication possible as the locus of the social meaning that hate/bias crimes supposedly express. Again, there is nothing nonsensical in such a proposal. Consider the often-used example of a novel typed by the five hundred thousandth monkey in the British Museum (who, like all the others, randomly strikes keys, but who, unlike all the others, lucks out in striking just those keys that produce not only syntactically correct sentences but also strings of them with theme, narrative consistency, interesting plot development, etc.). Such a production has meaning even if no one reads it. Its meaning lies in its conformity to the semantic, pragmatic, and stylistic conventions surrounding novelistic utterances in English. The text itself in this sense “has meaning” even though there is no novelist who meant anything by it and even though there is no

57
audience who understood anything because of it.\textsuperscript{86} The meaning such conventions create would probably be what a novelist meant if he wrote such things; it would probably also be what an audience understood if it read such things; but the text has meaning even without such an author or audience.

For our purposes, the equation of conventionalist meaning with the “social meaning” of crimes would parse like this: If a criminal action is accompanied by behavioral nuances that would be the nuances of one who hates or is prejudiced against his victim – words for sure, but physical acts of certain kinds, too – then the crime will \textit{look like} a hate- or bias-motivated crime. It will look like a hate/bias-motivated crime by the conformity of these behavioral nuances to the conventions by which hatred and prejudice is perceived. And according to this conception of social meaning, appearances are everything. That is, a crime that \textit{looks like} a hate crime will, on this view, \textit{be} a hate crime, i.e., a crime expressing a hateful social meaning.

With the conventionalist sense of “social meaning” being thus unproblematic, let us ask our normative question again: why would conformity (of a criminal action) to the conventions of hateful expression make that criminal action deserving of greater punishment? Remember, we cannot repair to the greater culpability arguably possessed by a criminal who experiences the emotion of group hatred or the stereotypical beliefs that spawn prejudice; nor may we repair to the harms that may be caused by beliefs about such subjective motivations. These are the speaker- and audience-centered construals of “social meaning” that we have already put aside. Rather, the question is why bare conformity of a criminal’s action to the conventions of hatred –

\textsuperscript{86} Such authorless, audience-less texts possessed of meaning are explored in Moore, \textit{Interpreting Interpretation}, in Moore, Educating Oneself in Public, \textit{supra} note 55.
the creation of an appearance of hatred, in other words – merits greater blameworthiness deserving of greater punishment.\footnote{This normative question is explored in Marcia Baron, \textit{The Moral Significance of How Things Seem}, 60 Md. L. Rev. 607 (2001).}

The only answer to this normative question that strikes us as at all plausible is one sounding in negligence. Actions that look for all the world like hate- or bias-motivated crimes run a substantial risk of causing the kinds of individual, group, and social harms catalogued before. Since such risk-taking is altogether without good reason – as its only justification is criminal, after all – it is an unreasonable risk to take. For this reason, those who commit crimes that appear to be motivated by hate or bias deserve greater punishment.

Now it is no doubt clear what we are going to say: Redundancy. The expressivist who tenders this conventionalist construal of “social meaning” gives us nothing more than was provided already by the non-expressivist who predicated greater blameworthiness on the allegedly greater harms caused by hate- and bias-motivated defendants. Not only did we find such claims deficient earlier, but the important point here is that expressivism (on this conventionalist construal) adds nothing to the case for hate/bias crime legislation that is not less confusingly put in non-expressivist terms.

The upshot is that the expressivist who wishes to justify hate/bias crime legislation on the expressivist ground that hate crimes send messages that demand contradiction by punishment must find some conception of a “message” that is “expressed” by a crime that is different than any of the foregoing. The expressivist must thus make sense of how a hate/bias crime can express a message that possesses a social meaning where that meaning is not identical to or given by: the intentions or other mental states of the offender; the beliefs or other mental states of
victims or of others; or the conventions affiliated with expressions of hateful emotions or bigoted beliefs.

Rather amazingly, most expressivists seem to think they can do exactly this. Thus, Elizabeth Anderson and Richard Pildes, in a leading article on legal expressivism generally (and not on the expressivist justification of hate/bias crime legislation, in particular), tell us that public meaning “need not be in the agent’s head, the recipient’s head, or even the heads of the general public.” They earlier give this example: “Musicians can play music that express sadness, without feeling sad themselves. The music they play need not express their (or anyone’s) sadness…”

A musical performance can certainly express sadness without the composer or the musicians themselves feeling sad. It can also express sadness without an audience on a given occasion being caused either to feel sad or to believe that the music is sad. Yet can it still express sadness if the music conforms to no musical conventions associated with the conveyance of a sad mood, so that it has no propensity whatsoever to elicit these cognitive or emotional experiences in audiences? Without actual mental states of authors or audience, and without any conventions abstracted from other authors and other audiences reacting to different, but relevantly similar, kinds of music, in what possible sense could music be sad or convey or express sadness?

---

89 Id. at 1508.
The possibilities here are not reassuring. Consider four of them. First, there is the evidential construal considered briefly by one perspicuous critic of expressivism, Matt Adler. On this construal, an event like an act of violence can have what the late Paul Grice called a natural meaning. “Clouds mean rain” was one of Grice’s examples, where “mean” obviously means no more than, “clouds are good evidence that it is going to rain.” Such natural meanings return us to examples like that earlier given: the rising of someone’s voice means anger. Anger is the meaning of the voice rising only in the sense that anger causes voices to rise and because of that, a voice rising is good evidence of anger. Yet this is to return us to an actor-centered, subjective interpretation of “social meaning,” and nothing more.

Second, there is the construal suggested by another perspicuous critic of expressivism, Simon Blackburn. Also struggling to come up with some sensible meaning to attribute to the expressivists’ idea of “social meaning,” Blackburn gives it what he calls a “credibility” construal: an action A expresses some meaning M when M is the only credible motivation an actor could rationally have had in doing A. Thus, in one of Blackburn’s examples, the flying of the Confederate flag over the government buildings in certain southern states expresses racism if the only credible reason a rational agent could have for flying such a flag is to celebrate the racist way of life formerly enjoyed in pre-Civil War days.

---

91 Grice, supra note 77, at ________.
92 Simon Blackburn, Group Minds and Expressive Harm, 60 Md. L. Rev. 467 (2001).
It is tempting to regard Blackburn’s credibility construal of social meaning as but another return to the subjective mental states of actors, the only novel twist being a stringent evidentiary requirement for when an actor can be thought to possess a mental state. Yet Blackburn plainly intends no such construal, for his aim is to provide a notion of social meaning that can apply to groups and institutions who in no obvious sense actually possess mental states. Blackburn’s is a constructivist construal of social meaning, one that is based on the mere appearance (as opposed to the reality) of their being a certain motivation. Such an appearance is generated by attributing rational beliefs and rational inferences to the agent (or entity) whose act is said to have social meaning and asking (in light of the causal and other properties of an action, and in light of the absence of any better means to achieve a certain state of affairs) what end could most rationally have been achieved by such a means.

Just how uniquely the credibility condition determines possible constructions of the motive(s) to be attributed to an action depends on how stringent the norms of rationality are taken to be. Make them very stringent and there may be but one rational end for a given action; make them very loose and many possible motives will exist for any action. Presumably we are to set the level of stringency for these norms of rationality at what we take to be the average information base, average inference drawing capacities, and average intelligibility of desires of humankind.

Blackburn’s construal of “social meaning” thus holds out the promise of giving the concept sufficient content that it avoids what Blackburn rightfully fears for it, namely, that “anything goes”\(^{93}\) for expressivists who assign social meanings to various actions. Yet

\(^{93}\) Id. at 479.
Blackburn’s construal renders social meaning unable to do the work assigned it by the expressivist justification of hate/bias crime legislation. Like all conventionalist construals of social meaning, Blackburn’s constructivist construal would make a defendant more blameworthy (and thus deserving of greater punishment) not because of any fact about him or his deed; rather, he would be subjected to increased punishment because of the appearance of there being a fact about him, namely, the appearance that he possessed a hateful or bigoted motivation for his crime (regardless of whether he in fact possessed such a motivation). We noted before that the appearance of a bad motive for an action can be a consequence of that action. Therefore, if the consequence is a bad one, the offender can be said to be guilty of wrongdoing beyond that perpetrated upon the principal victim. Three points must be made, however. First, usually the creation of an appearance of a bad motivation is not itself a wrong; our actual motives may matter to our blameworthiness, but how they look to others is not usually thought to be our responsibility. Others are responsible for their own beliefs and if they get things wrong, that is usually their problem and theirs alone. Second, when we are plausibly subject to an obligation not to create an appearance of improper motivation by our actions, that obligation appears to be a function of the bad consequences that are caused by the bad appearance. These are the consequences we have catalogued before: psychic distress to a victim, psychic trauma to those who share the victim’s characteristics, general social unrest, etc. As such, the appearance of bad motive (the constructed version of the social meaning of the offender’s action) does no more work to justify the greater punishment of the offender than is already done by the harm-based arguments for greater wrongdoing already examined in Part I. Third, in the conceivable but
probably non-existent case of non-consequentialist “mimetic” wrongdoing, the moral significance of a false appearance of bad motive surely is a lightweight in the justification of severely increased punishment. In what appears to be a racially motivated beating (where race in fact has nothing to do with the beating), the great wrong is in the violation of the victim’s bodily integrity and the great culpability is the offender’s intention to do just that. It may be (as we shall explore in the next Part) that an actual motive of racial hatred would increase the offender’s culpability. But the wrong of creating a false appearance of racial animus and the culpability of being negligent with respect to such a false appearance, seem trivial in contrast. Such trivial increments of wrongdoing and culpability hardly seem capable of justifying the severe sentencing enhancements of standard hate/bias crime legislation. Compared to enhancing the punishment of those who really are bigots, for example, punishing those who appear to be bigots is small potatoes.

A third construal of social meaning is offered up by Deborah Hellman in her expressivist approach to equal protection analysis in constitutional law. Hellman appeals to audience beliefs in her construal, but they are not the actual beliefs of crime victims or others; rather, we are to seek the hypothetical beliefs that a populace would have if they were to dialogue a bit in a Habermasian Ideal Speech situation: “[T]he expressive dimension of a law or policy is best

---


95 Id.
understood as the meaning that we would arrive at if we were to discuss the interpretive question together under fair conditions.”

It is unclear how an expressivist who seeks to justify hate/bias crime legislation would flesh out this suggestion. What would be the “fair conditions” for a discussion about the “interpretive question” of the social meaning of some criminal’s action? But for present purposes it does not matter how this suggestion would be fleshed out. If the idealized epistemic procedure (whatever it is) is followed, and if it yields the social meaning “racist,” that surely can have nothing to do with the desert of an offender. For the outputs of such idealized epistemic procedures cannot even claim the relevance of an appearance of bad motivation; such outputs can only yield the conclusion that there would be such an appearance in the idealized situation imagined. It is morally unintelligible how one could conclude that those whose crimes would be viewed as racist if people engaged in certain idealized dialogues are therefore in fact guilty of greater wrongdoing.

A fourth construal of the expressivists’ key concept of social meaning is offered by another critic of expressivism, Steven Smith. Smith also engages in an extensive search for a conception of “social meaning” that both is coherent and can do the normative work assigned to it by expressivists. In exasperation Smith concludes that “expressivist scholars seem to suggest that laws or actions can just have ‘objective’ meanings, period – meanings that need not be

---


meanings of or to anyone.” 98 Smith concludes that “this assertion simply renders the notion of ‘meaning’ unintelligible. . .”.99

Our own brief perusal of the expressivist literature confirms Smith’s diagnosis. Those expressivists who wish to free themselves from the constraints on what social meanings can be – constraints framed in terms of intentions, beliefs, or conventions – do seem to yearn for the existence of meaning as a metaphysically autonomous thing. Ironically enough, the most forthright example of this longing for a “just there” view of meaning is Smith’s commentator, Ed Baker.100 For Baker has discovered a “missing realm” of existence that solves all of these problems. There is, Baker assures us, both the “material realm” of the natural world and the “subjective realm” of the mind; but there is, in addition, a “third realm,” what Baker dubs the “social realm.”101 This solves the conceptual worry because “meaning exists here, in this social realm, not as ‘mental events’ in the heads of either creators or perceivers.”102 And this construal of social meaning also solves our normative worries, for Baker claims “that a proper normative concern focuses on ‘meanings’ that exist in the social realm . . . .”103 Baker concludes that “an understanding of this ‘social’ realm is absolutely fundamental not only to an understanding of meaning and interpretation, but of the social sciences generally.”104

---

98 Id. at 562.
99 Id. at 563.
100 C. Edwin Baker, Injustice and the Normative Nature of Meaning, 60 Md. L. Rev. 578 (2001)
101 Id. at 584-585.
102 Id. at 586.
103 Id. at 585.
104 Id. at 584 n. 9.
In making these extraordinary metaphysical assertions Baker quite explicitly puts himself in the tradition to which they belong.\textsuperscript{105} This is the hermeneutic tradition begun by Dilthey in Nineteenth Century theology, a tradition according to which the meaningful “belongs to a level fundamentally different from the physical.”\textsuperscript{106} Investigation of this “level” or “realm” has its own goal (understanding, not knowledge) and its own investigative method (interpretation, not description and explanation). This realm also has its own criteria of success: although “‘objectivity’ means something quite different in the \textit{geisteswissenschaften} compared with the natural sciences,”\textsuperscript{107} nonetheless an understanding or an interpretation can be correct in the social realm when it “correspond(s) fully to the meaning underlying the text as an objectification of mind.”\textsuperscript{108}

As one of us has charted in detail elsewhere, “special realms” in general, and this special realm in particular, are the last refuges for a theory in trouble.\textsuperscript{109} Metaphysical divides between the natural and the moral, between the physical and the mental, or between the individual and the social, make for queer entities within the special realm, queer relations between entities in different realms, and a queer epistemology about how one gains knowledge of what goes on in

\textsuperscript{105} Id. at 591, n. 18, 592.

\textsuperscript{106} This tradition is explored in Moore, \textit{The Interpretive Turn in Modern Theory: A Turn for the Worse?}, in Moore, Educating Oneself in Public, \textit{supra} note 55, at 384-387.

\textsuperscript{107} Emilio Betti, \textit{Hermeneutics as the General Methodology of the Geisteswissenschaften} in Contemporary Hermeneutics, 63 (J. Bleicher ed., Routledge, 1980).

\textsuperscript{108} Id. at 79.

\textsuperscript{109} Moore, \textit{Interpretive Turn, supra} note 106.
the special realm. And even if we pass by all of these well worn objections to Baker’s kind of pluralistic metaphysics, there is also Simon Blackburn’s worry here, namely, that we have now entered the “hermeneutic desert,” a “landscape where anything goes.” With friends like Baker, expressivists need no enemies.

It is clear to us that expressivism cannot deliver up a concept of social meaning that is: (1) coherent and determinate in its implications for particular cases; (2) capable of doing the normative work assigned to it in the justification of hate/bias crime legislation (namely, to show greater wrongdoing when the social meaning of a defendant’s crime is racist, homophobic, chauvenistic, etc.); and (3) non-redundant of non-expressivist accounts of greater wrongdoing predicated on claims of exacerbated harms to the interests of victims and others. The first version of the expressivist analysis of hate/bias crime legislation (the “moderate” version) thus falls away.

C. The Expression of Hate and Prejudice as Triggers for Enhanced Punishment Under an Expressivist Theory of Punishment

The “radical” interpretation of the expressivist thesis relieves the expressivist of the burden of showing how the expression of hate/bias (in some sense of “expression”) makes a hate/bias crime deserving of enhanced punishment as against the base crime. For this second

---

111 Blackburn, *supra* note 92, at 482.
112 *Id.* at 479.
version of expressivism replaces the traditional justifications of punishment (predicated on desert and/or crime-prevention) with an altogether new ideal. On this theory, punishment itself sends a message, and the justification for delivering this message resides in the power of punishment to contradict or counteract the messages of hate and prejudice conveyed by hate/bias crimes.

For this branch of the theory, any coherent, determinate notion of social meaning will do. That is, one can take the message expressed by the criminal to be: (1) the actual motive of the criminal; (2) the actual beliefs about the criminal’s motive, by the victim or others; or (3) the conventions surrounding rational goal-seeking or rational emotionality that are responsible for giving some criminal deeds the appearance of having been badly motivated. Any of these can be the social meaning of the criminal’s act, because on this analysis that meaning needs neither to make the act a greater wrong nor to make it punishable on grounds that are not redundant with the harm-based accounts explored in Part I.

Still, it would be helpful in focusing the discussion that follows if we were to pick one meaning of “social meaning.” Let us settle on the most obvious meaning: a criminal act expresses a social message of hatred or bias if the criminal is actually motivated by hatred or bias. The enhanced penalties of hate/bias crime legislation will then be said to express social disapproval of the hate/bias crime perpetrator’s message if that is what legislators and judges actually intend by their application.

Of all the purported justifications of hate/bias crime legislation that we have examined, this “send a message to counteract the message of hate-motivated offenders” seems to be the
single most persuasive arguments to proponents of such legislation. We should thus examine it with some care. There are three questions to ask about this argument. Exploring them will both clarify the radical expressivist argument and show it to be a poor justification for hate/bias crime legislation.

1. Why Is Sending a Message to Criminals Good?

Expressivists of the stripe here considered have to think that it is good for society to express its disapproval of hate- and bias- motivated crimes—indeed, so good that it can justify the intentional infliction of suffering that is punishment. Our first question is why one should think this.

There are some easy answers that are not open to expressivists. These are answers that derive from a utilitarian theory of punishment which takes the expression of societal outrage as an *instrumental* good in the service of other, intrinsic goods. Thus, a utilitarian might plausibly suppose that the denunciation of crimes, criminals, their motives, etc., will: (1) vent a society’s vengeful emotions in a way that prevents some members of that society from taking the law into their own hands; (2) reinforce the values of law-abiding citizens in a way that keeps them law-abiding; (3) shame others who are less law-abiding in their inclinations from breaking the law; (4) educate the populace in the principles of morality that keep most people from being harmed

---

by others; (5) satisfy what most people most want; (6) maintain a sense of social cohesion by giving members of society an identity defined by a set of shared values; etc.

The radical expressivist cannot appeal to this familiar list of goods without reducing expressivism to a mere means of achieving a utilitarian program. We were promised more. We were promised a theory of punishment independent of the traditional theories of punishment. If radical expressivists are to deliver on this promise, the expression of disapproval by criminal punishment must be an *intrinsic* good, not merely an instrumental good in the service of utilitarian goals.

Expressivists need their theory to be independent of traditional utilitarian and retributive theories because they claim for it an ability to justify something–namely, hate/bias crime legislation–in a way that the traditional theories cannot. Suppose that hate/bias crime legislation cannot be justified on utilitarian grounds as an efficient means of achieving crime prevention, and suppose that it cannot be justified on retributivist grounds as a proportional response to increased desert (suppositions still to be fully explored in the remaining Parts of the paper); still, the claim is, hate/bias crime legislation can be justified by this new expressivist theory of punishment, for it relies neither on claims of crime prevention nor on claims of desert in advancing its justifications of penal institutions.

For this reason, an expressivist cannot pretend to be a kind of retributivist any more than he can pretend to be a kind of utilitarian. One version of this pretend-retributivism is the relativist’s version. We may, for example, fairly characterize Lord Devlin’s well known tirade against homosexuals to rest on a relativistic view of morality. Whatever a society most strongly disapproves of – measured by the fevered pitch of intolerance, indignation and disgust – is what
is wrong in that society. On this view there can be no separation between just deserts and popular, deeply felt beliefs about just deserts. If a society feels more strongly towards hate/bias crimes than toward other crimes, then hate- and bias-motivated criminals deserve more punishment than other criminals.

No retributivist can accept Lord Devlin’s kind of relativistic meta-ethics and remain a retributivist. Just deserts cannot be collapsed into popular attitudes about just deserts and remain a trigger for justified punishment. One thus cannot give a retributivist’s justification of hate/bias crime legislation on the grounds that greater social hatred of the message of a hate/bias crime constitutes a basis for attributing to the perpetrator “desert” of greater punishment, for the supposed desert is only a belief about desert and not desert itself.

Similarly, the expressivist cannot pretend to be a retributivist by confining his use of the expressive function of punishment to the interstices within the vague injunctions of desert. One cannot urge, as does Frederick Lawrence, that even if the expressive value of punishment in no sense justifies why we are entitled to punish in general, it nonetheless “must inform our decisions about the nature of that punishment” and judges must measure out the amount of punishment by the degree of offensiveness of the offender’s conduct to society. Nor can one urge, as Dan Kahan does, that it is possible for a retributivist “to use the expressive view of punishment to inform desert. . . . The proper retributivist punishment is the one that appropriately expresses condemnation and reaffirms the values that the wrongdoer denies.” If giving just deserts justifies punishing at all, it justifies what kind of punishment an offender should receive, and

114 Devlin, supra note 46.
115 Lawrence, Punishing Hate, supra note 3, at 167.
how much. There is no room on retributivist grounds for something other than desert – the expression of societal outrage – to intrude. Moreover, if something else, such as the expression of social outrage, is limited in its intrusions to cases in which desert is indeterminate about details such as the exact kind and amount of punishment owing, then the something else is doing all of the justifying work about these details. And then our question recurs: what makes expressing a society’s disapproval of the motive of a criminal’s act intrinsically good? It cannot be because of desert, so what is it?

We can think of no plausible answer to this question. Some institutional arrangements can plausibly be argued to be intrinsically good, such as those educational institutions that come as close to giving equality of opportunity as is humanly possible. But using the criminal law to express popular attitudes does not seem at all like this. To regard the expression of social attitudes as intrinsically good smacks of a kind of collective first amendment right, a right justified by the intrinsic goodness of “self” expression. Yet when the self is not an individual, such an intrinsic good is not at all plausible. It seems a kind of category mistake to liken the collective expression behind a criminal law to the individual expression in a work of art.

It might be urged that at best we have reached a standoff on this issue. For if a good is said to be intrinsically good, what argument is possible either for or against it?\textsuperscript{117} We say no, you say yes, and there is no more to be said about it. Yet intrinsicality is not an epistemic property; it is a metaphysical property. Because something is intrinsically good does not mean that no reasons can be adduced making it rational to believe that that thing is intrinsically good. It is indeed true that one cannot show something to be intrinsically good by showing how it

\textsuperscript{117} On how one justifies first principles, and how such principles differ from intrinsic goods, see Moore, \textit{Placing Blame, supra} note 5, at 159-188.
contributes to something else that is good. But this is also true for categorical obligations, which one cannot justify by citing good consequences. But non-consequentialist justifications are possible, in terms of a mix of more particular judgments and more general principles one provisionally accepts as true. Thus, it is false that nothing can be said to justify a belief in the intrinsic goodness of an act, an institution, or a state of affairs. The problem for the radical expressivist is that nothing comes to mind by which to justify a belief in the intrinsic goodness of an institution that expresses social revulsion or disapproval.

2. Is Sending a Message So Good That It Justifies Punishing an Offender More Than is Deserved?

Suppose that we find some basis for concluding that sending messages that disapprove of the messages of hate/bias crime perpetrators is good. The next question is whether such message-sending is so good that it could justify punishing those guilty of hate/bias crimes beyond the punishment that they deserve. We reach this question because on this more radical branch of expressivist theory we are assuming we have a new theory that can justify punishment even when desert-oriented retributivist and mixed theories cannot.

When dealing with the utilitarian theory of punishment in philosophy, it is standard fare to argue against that theory by way of counterexamples.118 Such counterexamples consist of cases in which there is a good utilitarian case for punishing some wholly innocent individual, and

118 As in Moore, Law and Psychiatry, supra note 51 at 238-240.
the question put to the utilitarian is whether he can stomach the intuitively unjust implications of his own theory.

Less widely recognized is the fact that such counterexamples can be constructed for any consequentialist theory, whether it be utilitarian or not. Consider a consequentialist-retributivist\textsuperscript{119} who seeks simultaneously to maximize the cases in which the guilty receive the punishment they deserve and the cases in which the innocent are free of punishment that they do not deserve. Such a theorist is subject to the same sorts of counterexamples as is a utilitarian. Let it be the case that by punishing one innocent person the retributivist-consequentialist could both convict more guilty people and free more innocent people. For example, suppose the punishment of one innocent person could yield testimony sufficient both to convict other guilty parties and to free innocent parties wrongly convicted of a crime. Logically the consequentialist-retributivist is committed to punishing the innocent in order to obtain the greater goods imagined.

What such a counterexample demonstrates to most people is that they are not consequentialists at all about punishment. Rather, they regard the obligation not to punish the innocent as a “side-constraint,” an “agent-relative” or “categorical” obligation. We are not justified in violating such obligations even when our doing so on one occasion would minimize other violations of that very same obligation in the future.

Now return to the proposed expressivist theory of punishment. Such a theory is fully subject to the kinds of counterexamples just imagined. Assume (arguendo to be sure) that it is very very good to express our condemnation of crime and criminals through punishment. Still,

\textsuperscript{119} On consequentialist retributivism, see Moore, Placing Blame, \textit{supra} note 5, at 155-159.
the attainment of that good cannot justify punishing the innocent, so long as one is persuaded by the above sorts of counter-examples to utilitarianism demonstrating that punishment must at least be side-constrained by the offender’s just deserts.

If we were expressivists about punishment we would respond to such an argument in the following way. The obligation to punish an offender only so much as he deserves does not operate as a side-constraint to attaining the good of expressing condemnation; it is internal to that good itself. That is, the good of expressing condemnation is a good at all only when the criminal being condemned fully deserves the punishment that expresses the condemnation. Because this is an intrinsic good, it is not possible to achieve such a good absent desert. It is possible, of course, to achieve the educational and other benefits caused by expressions of condemnation even when the person who is condemned is innocent; but these benefits are not what makes expression of condemnation intrinsically good.

Whether this imagined response is true of the expressivist’s theory depends entirely on how he construes his expressivism. If his theory is consequentialist, then this response will be false. For we can easily construct situations in which we could maximize the expressions of condemnation for those who deserve it by expressing condemnation of one who does not deserve it. The consequentialist-expressivist differs not a whit in this vulnerability from the retributivist-consequentialist earlier imagined.

So to make the response above true, the expressivist must construe his theory of punishment as deontological in character. We are categorically obligated not to express condemnation of those who do not deserve it even when such an expression would maximize the expression of condemnation of those who do deserve it. Now the expressivist’s response is true.
Unfortunately, however, such a response completely trivializes the theory. For now the expressivist theory of punishment is incapable of justifying any punishment beyond what is deserved. In particular, the theory cannot justify hefty sentence enhancements for crimes committed with hateful or bigoted motivations if those motivations do not increase the blameworthiness (desert) of the offenders who possess them. The expressivist is thus thrust back to the other, less radical branch of the theory—the branch according to which hateful and biased motivations do increase offenders’ blameworthiness, and thus appropriately subject offenders to greater punishment under standard retributivist or mixed theories. Yet, as we earlier endeavored to show, that branch of the theory is equally indefensible.

Since we take all of this to be relatively obvious, the question remains as to why so many proponents of hate/bias crime legislation regard the need to “send a message” to be a sufficient justification for the law. We think it is because such proponents are focusing only on the passage of the bills enacting hate/bias crime legislation. It is this legislative act that sends the desired message. Forgotten is the obvious fact that when one uses the criminal law as a medium for sending a message, one then has to punish those who do not get the message. If such persons do not deserve such punishment, then the cost of sending messages by means of the criminal law is rank injustice.

It is true that if everyone got the message, then no one would actually suffer undeserved punishment. The legislature would only threaten an undeserved punishment, and one might well regard it as morally unproblematic to threaten to do what one is not in fact morally permitted to do. The problem, of course, is that prohibiting hate-and bias-motivated crime will not, by itself,
successfully stop it. The threat of punishment in excess of what is deserved will thus have to be carried out.

The only way to avoid this unacceptable cost is by failing to enforce hate/bias crime laws. With sufficient acoustic separation, a legislature could express its condemnation of racism, sexism, and bigotry by passing hate/bias crime legislation and yet avoid the cost to justice by sending a different message to prosecutors: “Hate/bias crime legislation is merely symbolic; it is not real criminal legislation, so do not prosecute.” Sara Beale, for example, has recently advanced a version of this proposal in defense of the proposed federal hate/bias crime bill, when she urges “a greater emphasis on the symbolism of the enactment of the law as compared to its enforcement. . . .” She finds that expressivism provides “a more satisfactory theoretical basis for the creation of broadly phrased federal hate crimes that will seldom be prosecuted.”

“Seldom,” of course is not never, so those who adopt Beale’s line have to be willing to put up with some undeserved punishment as a result of the enactment of hate/bias crime legislation. Moreover, the only reason Beale can plausibly predict that federal hate/bias crime prosecutions will seldom occur is because of deference by federal prosecutors to state hate/bias crime prosecutions. Unless state prosecutors also understand that they are not to prosecute under state hate/bias crime laws, undeserved punishment will be an often-experienced cost of hate/bias crime legislation.

---

121 Beale, supra note 113, at 1265.
122 Id. at 1256.
Finally, there are well known costs to “symbolic,” i.e., unenforced criminal laws. One of those costs is a cost to the very message that is sent by those laws. Acoustic separation breaks down the moment citizens eavesdrop on the secondary message to prosecutors and judges. If a crime is never prosecuted, citizens come to regard it as not really a crime at all, and its message of condemnation is therefore lost. Thus, any attempt to employ a strategy of non-enforcement as a means of minimizing the costs to justice imposed by hate/bias crime legislation takes away the gain that was to justify imposing the costs to justice to begin with.

3. Is Sending a Message So Good That It Justifies Punishing an Offender Less Than is Deserved?

The other kind of counterexample standardly advanced against a utilitarian theory of punishment plays on the observation that sometimes punishment of the guilty is not necessary to achieve the utilitarian goals of crime-prevention, the preservation of social cohesion, etc. Punishment in such cases is fully deserved, however. So utilitarians are committed to letting the guilty go free in some circumstances, no matter how heinous their crimes and no matter how great their desert.

Counterexamples that demonstrate this point can be advanced against any form of consequentialist theory of punishment. A consequentialist-retributivist, for example, is committed to freeing the guilty whenever it will produce more punishments of more guilty

---

124 Moore, Law and Psychiatry, supra note 5, at 240-242.
persons—e.g., by giving immunity to an offender in order to obtain testimony against, and conviction of, a greater number of other criminals (or another more blameworthy criminal).

The particular form that this objection takes to an expressivist theory of punishment begins with the observation that we could express our condemnation of criminals without actually punishing them.\textsuperscript{125} We could obtain the good sought by the expressivist theory—denouncing what deserves denouncing—without in any other way making those denounced suffer. The publication of offenders’ names and offenses, the production of plays with great choruses denouncing named criminals, the use of scarlet letters, the imposition of punitive damages in tort, and the like, come to mind as vehicles of public condemnation that involve little of the suffering that we think of as the essence of punishment. If condemnation can sometimes better be achieved through means other than punishment, then on an expressivist theory of punishment (as on a utilitarian theory of punishment) desert drops out as a sufficient condition of punishment. Just because an offender richly deserves to be punished is not reason enough to punish him if he can be condemned without such punishment.

In response to this old objection, expressivists repair to “what convention and form contribute to meaning.”\textsuperscript{126} Conceding that “in some societies, and even in our own at an earlier time, public denunciation by itself might have been sufficient to convey condemnation of a wrongdoer,”\textsuperscript{127} in our society now only the imposition of suffering on an offender can express

\begin{footnotes}
\footnotetext[125]{H.L.A. Hart, Law, Liberty, and Morality 65-66 (Stanford, 1963).}
\footnotetext[126]{Kahan, Alternative Sanctions, supra note 71, at 600.}
\footnotetext[127]{Id.}
\end{footnotes}
our condemnation of him and his action: “the way for society to show that it takes rape seriously, and to show that it genuinely condemns a particular rapist, is to make him suffer . . .”\textsuperscript{128}

Even supposing that this bit of sociology is true (as it well might be), such a response hardly meets the challenge posed by the counterexample above. Such a counterexample is presented as a thought experiment. Such a thought experiment is designed to let those who run them test what it is they really believe about a theory by discovering whether they accept its implications in all imaginable cases. That in our actual world there is a contingent association between condemnation and suffering is irrelevant. The question an expressivist should be answering is whether in all possible worlds she can stomach the implications of her own theory. Can she accept the implications of her theory in the case of a brutal rapist who fully deserves punishment within a society that is not like ours—that is, within a society in which verbal condemnation is understood as sincere and deeply meant in the absence of imposed suffering?

Furthermore, like the utilitarian, the expressivist is faced with the possibility that a society might (and ought to) merely pretend to punish whenever such a pretense can be carried out without detection. Why not institute a brutal rapist protection program, which gives rapists a new identity in a far away place, while using the conventions of condemnation that exist in our society to make a great show of (seemingly) punishing such long-gone offenders?

Of course, like the utilitarian, the expressivist can protest that there is an ineliminable risk that such dissembled punishments will be detected. Yet such risks only give us a discount rate. Surely if it is so good to achieve public condemnation of criminals and so bad to make people suffer, then we should take some risks whenever we can achieve the good without the bad.

\textsuperscript{128} \textit{Id.} at 600-601.
For each of these three reasons we view the radical version of the expressivist thesis as an unacceptable justification of hate/bias crime legislation. There is no “expressivist theory of punishment” that is even half-way respectable as a theory of punishment. Even Joel Feinberg, whose early article on “the expressive function of punishment” is often cited by present day expressivists, always held that expression of condemnation is at most an incidental function of punishment, and is no part of what properly can be called a theory of punishment, viz, a theory that justifies the institution. Without such a theory, expressivists are thrust back to the first version of their defense of hate/bias crime legislation—the version under which they must show how an offender’s expression of hatred or prejudice by itself increases the wrongness of his crime. And as we have seen, this is to return from the frying pan to the fire.

III. THE CULPABILITY ANALYSIS

While advocates of hate/bias crime legislation frequently reach to the arguments that we have canvassed in the previous two Parts, the language of typical hate/bias crime statutes favors the view that hate and bias toward a victim because of her membership in a particular group constitute uniquely culpable mens rea. While the actus reus of a crime specifies a particular

130 E.g., Kahan, *Alternative Sanctions*, supra note 71, at 595 n. 11.
131 Personal Communication of Joel Feinberg to Michael Moore.
132 See e.g., New Jersey Code of Criminal Justice, 2C: 33-4 (a person commits a crime of the fourth degree if in committing an offense under this section, he acted with a purpose to intimidate an individual or group of individuals because of race, color, religion, gender, handicap, sexual orientation or ethnicity.); 18 Pa. C.S.A. sec. 2710 ([M]alicious intention
kind of prohibited wrongdoing, the mens rea of a crime specifies the culpability with which a
defendant must have acted in order to be punished for that wrongdoing. On the third justification
of hate/bias crime legislation, then, hate and bias constitute particularly culpable subjective
mental states on the part of defendants who engage in traditionally prohibited wrongdoing.
According to this third justification, the fact that a defendant is motivated to harm his victim as a
result of hatred for the victim’s race or religion makes the defendant more culpable, as a moral
matter, for the harm he perpetrates than he would be if he purposefully perpetrated the same
harm for a different reason. Inasmuch as a defendant who is more culpable for a wrong is more
to blame for its commission than is another who commits the same wrong less culpably (so that
the premeditated murderer is more to blame than the negligent killer), one who commits a
hate/bias crime is more blameworthy than an otherwise motivated offender, and hence, on
traditional retributivist or mixed theory principles, is deserving of greater punishment.

If, in keeping with this third “culpability analysis,” hate and bias are construed as mens
rea requirements, rather than actus reus requirements, then hate/bias crimes differ from all other
sorts of crimes in very significant respects. As we shall argue, hate/bias crimes are concerned
with defendants’ motivations for action in a way that no other crimes have ever been concerned.
This is because the motivations with which they are concerned are emotional states or
dispositional beliefs that attend actions, rather than, say, future states of affairs to which actions
are instrumental means. These emotional states and dispositional beliefs constitute standing
character traits rather than occurrent mental states (such as intentions, purposes, choices, etc.).

means the intention to commit any act, the commission of which is a necessary element of any
offense referred to in subsection (a) motivated by hatred toward the race, color, religion or
national origin of another individual or group of individuals.)
Thus, the additional penalties that are imposed on defendants who are found guilty of hate/bias crimes constitute punishments for (bad) character.

In the first section of this Part, we shall work through the particulars of the argument that if hate/bias crimes punish defendants for their culpability, then they punish them for bad character. In the second section, we shall explore the moral and political significance of criminalizing bad character.

A. Group Hatreds and Biases as Character Traits

If hate/bias constitutes a subjective mental state with which a defendant must act in order to be liable for a hate/bias crime, then the mens rea of hate/bias crimes is unlike the mens rea required for any other crime. The standard method of grading culpability in criminal law has long been indifferent (except evidentially) to considerations of motivation. Most criminal statutes have required fact-finders simply to answer one of the following questions regarding the mental state of a defendant who has performed the actus reus of a crime, and has thus brought about a legally-prohibited result: Did the defendant have as his purpose the legally prohibited

---

133 The exceptions are discussed below. As will become clear, even doctrines that commit fact-finders to more extensive mens rea inquiries do not compel them to examine the sorts of motivational states with which hate/bias crimes are concerned.

134 The Model Code implicitly distinguishes between crimes that prohibit certain results and crimes that prohibit conduct of a certain nature. Model Penal Code sec. 2.02 (2)(a)(i) and (ii) (Proposed Official Draft 1962). It does this because its drafters thought that some prohibited conduct is without consequences. This is false, but in any event, it is a nicety that need not concern us here. Let us just stipulate for the sake of this discussion that whenever we talk of legally prohibited results we are referencing both what the Model Penal Code would term legally prohibited consequences and what it would term conduct of a legally prohibited nature.
result? Did the defendant at least know that he would cause such a result? Was he consciously aware of a substantial and unjustifiable risk that he would cause the prohibited result--i.e., was he reckless in bringing it about? Or should he have been consciously aware of a substantial and unjustifiable risk that he would cause such a result--i.e., was he negligent with regard to causing that result? None of these inquiries require consideration of the reasons for which the defendant acted or the background motivations with which he performed his actions (except, of course, evidentially—that is, such that, by virtue of his motivation, the defendant can be found to have possessed the requisite purpose, knowledge, recklessness, or negligence required for liability). In contrast, inasmuch as hate/bias crimes require fact-finders to engage in such a motivational inquiry, their enactment breaks whole new ground in the development of criminal law doctrine.

Some will surely object that this characterization of the mens rea of hate/bias crimes overstates their novelty. There are, after all, a number of traditional doctrines that appear to require inquiries into defendants’ motivational states, e.g., the provocation/passion,

---

135 This is the question that a court is required to ask when the crime with which the defendant has been charged is, in common law parlance, a specific intent crime.

136 This is the question that a court is required to ask when the crime with which the defendant has been charged is, in common law terms, a general intent crime.

137 This is the question that a court is required to ask when the crime with which the defendant has been charged is, in common law vernacular, a crime requiring malice.

138 For the most influential recommendation that every criminal statute be interpreted to require a finding of purpose, knowledge, recklessness, or negligence with regard to the nature of conduct and consequences prohibited by the statute, see Model Penal Code sec. 2.02. The Model Penal Code mens rea doctrines, are, of course, textured, in that they alter the meaning of the term purpose as between actus reus requirements that concern the nature of the defendant s conduct, the consequences caused by the defendant s conduct, and the circumstances within which the defendant acts. These distinctions are of no concern to us here, because none of them invite motivational analyses of the sort required by hate/bias crimes.
premeditation/deliberation, and specific intent doctrines. Let us thus work through these other doctrines so as to be clear about how the mental state inquiries that they require differ from the mens rea determinations required by hate/bias crimes. Such an exercise will expose the doctrinal novelty of hate/bias crimes and reveal the intrinsic nature of hate and bias as mens rea requirements.

First, there are clearly defense doctrines that require courts to consider good, or at least exculpatory, motivations as bases for reducing or altogether suspending penalties for prima facie wrongdoing. For example, if a defendant commits a prima facie wrong because he reasonably believes that it is the lesser of two evils, then he may invoke the justification of necessity. If he commits such a wrong because he reasonably believes that it is a necessary means of defending a third party against a culpable aggressor, then he may invoke the justification of third-party self-defense. And if a defendant kills as a result of being provoked into a passion in circumstances in which a reasonable person might become similarly impassioned, then the (common law) provocation/passion doctrine reduces his liability from murder to voluntary manslaughter. So certain good, or at least exculpatory, motivations are made material by traditional defenses (and

---

139 For a critique of the law’s willingness to extend justifications to defendants if and only if they both reasonably believe in the existence of justifying circumstances and act because of such circumstances, see Heidi M. Hurd, *Justification and Excuse, Wrongdoing and Culpability*, 74 Notre Dame Law Review ____, ____ (1999).

140 Inasmuch as the provocation/passion doctrine is concerned with an emotional state within which a defendant acts, it is more like the mens rea doctrine of hate/bias crimes than is any other doctrine in the criminal law. However, inasmuch as the provocation/passion doctrine is an exculpatory doctrine, while the mens rea doctrine of hate/bias crimes is an inculpatory doctrine, there is ultimately little analogy between the two, and the former is hardly a precedent for the latter.
partial defenses) in ways that fully or partially exonerate persons for well-motivated or understandably ill-motivated conduct.

Unlike these doctrines, however, hate/bias crimes require courts to consider bad, or inculpatory, motivations as bases for enhancing penalties for wrongdoing. The doctrines that exculpate defendants from liability for well-motivated (or understandably ill-motivated) conduct thus provide no precedent for the new mens rea doctrines created by hate/bias crimes. Worse yet, in the case of the provocation/passion doctrine, hatred, anger, rage, and the like are thought to mitigate blameworthiness because they are thought to suspend the sort of reasoned judgment that is required for full responsibility. It would be flat out contradictory to point to the provocation/passion doctrine as a precedent for hate/bias crime legislation, for it would be to say that the very same condition that exculpates also inculpates.

There are, however, several other traditional doctrines that, at first glance, might appear to enhance a defendant’s punishment upon a finding that the defendant harbored particularly vicious reasons for action. Consider, first, the premeditation/deliberation doctrine. In a number of jurisdictions a defendant who premeditates and deliberates about a killing is eligible for more severe punishment than is a defendant who kills purposefully, but without premeditation and deliberation. Does an inquiry concerning whether a defendant premeditated and deliberated about a killing constitute the same sort of motivational inquiry into which fact-finders must delve in order to assess hate/bias crime liability? The answer is, no. To determine that a defendant premeditated and deliberated about a killing simply requires a fact-finder to find that the defendant formed the purpose to kill and contemplated the means by which to kill

141 We are grateful to Edward Rock for raising this doctrine as a possible precedent.
temporally in advance of the killing. In making this determination it may be evidentially valuable to discover that the defendant hated the victim or had a motive to kill the victim, since if he did, he may well have thought about killing the victim. But, as a matter of law, to find that a defendant premeditated and deliberated about a killing does not require the fact-finder to discover the reasons for which the defendant intended to kill and deliberated about its means, nor does it require the fact-finder to determine the emotional states within which the defendant formed the purpose to kill, considered how to kill, and killed.

Hate/bias crime liability, in contrast, commits fact-finders to assessing not just how a defendant came to do what he did, but why. By requiring a finding that the defendant harmed his victim as a result of hating or being prejudiced against a group of which she is a member, hate/bias crimes force fact-finders to assess not just whether the defendant formed a purpose to harm his victim, nor even just whether the defendant calculated about the means by which to do his victim harm. They further require fact-finders to assess the reasons for which the defendant harmed his victim and the emotional state within which he acted when perpetrating the harm. These inquiries thus far surpass those required by the premeditation/deliberation doctrine in homicide law.

As a third possible precedent for hate/bias crime doctrine, consider the category of specific intent crimes.\textsuperscript{142} A specific intent crime is a crime that requires a defendant to do a prohibited action with some \textit{further purpose} (beyond the purpose to do the prohibited act). Thus, for example, burglary is a specific intent crime because the defendant must have purposefully

\textsuperscript{142} See Paul H. Robinson, \textit{Hate Crimes: Crimes of Motive, Character, or Group Terror?}, Annual Survey of American Law 1992-93 (1992), pp. 606-09 (defending the claim that specific intent crimes provide a precedent for hate/bias crimes, but arguing that, notwithstanding these precedents, hate/bias crimes may not be the best means of accomplishing their own goals).
broken and entered into a building with some further intention, say, to steal, rape, or kill.

Attempted murder is a specific intent crime because a defendant must have purposefully taken a substantial step toward killing a person (e.g., taking aim at him) with the further intention of bringing about that person’s death. Assault with intent to rape (or kill or steal) is a specific intent crime a defendant must have purposefully inflicted a harmful or offensive contact on a person with the further intention of having sexual intercourse with that person.

Specific intent crimes like these all appear to have motivational mens rea requirements. “Indeed,” as Paul Robinson argues, “every time an offense definition contains the phrase ‘with the purpose to . . .’, the law takes as an offense element the actor’s motive, the cause of his or her act.” To assess liability, fact-finders must ask why a defendant did what he did, not just whether he did what he did purposefully. If a defendant was not motivated to break and enter into another’s building by a desire to commit a felony, then that defendant cannot be held liable for burglary. If a defendant was not motivated to pull the trigger of his gun by a desire to kill his victim, then that defendant cannot be held liable for attempted murder when he shoots a passing pigeon instead. Inasmuch as such crimes appear similar to hate/bias crimes in requiring fact-finders to seek out defendants’ motivations, they appear to falsify the claim that the enactment of hate/bias crimes constitutes a radical departure from, rather than extension of, traditional mens rea categories.

---

143 Robinson, id. at 606-07. For a similar defense of the claim that hate/bias crimes have precedents in specific intent crimes, see Jeffrie G. Murphy, Bias Crimes: What Do Haters Deserve?, 11 Crim. Just. Ethics 20, 21 (1992). For more general indictments of the over-used and over-broad slogan that motives are irrelevant to criminal liability, see Douglas Husak, Philosophy of the Criminal Law, chap. 5 (1987); Douglas Husak, Motive and Criminal Liability, 8 Crim. Just. Ethics 3 (1989).
But despite initial appearances, there are very important differences between the motivational mens rea of specific intent crimes and that of hate/bias crimes. Specific intentions have as their objects future states of affairs—states of affairs that constitute reasons for action. In contrast, hatred constitutes an emotional state within which an actor acts, while prejudice constitutes a disposition to make judgments about others and to act towards them on the basis of false beliefs about them.

By talking simply about targeting defendants’ motivations for action, we thus run the risk of obfuscating the distinctions between three sorts of motivations: desires, passions, and dispositional beliefs. These distinctions are crucial to differentiating between the motivations relevant to specific intent crimes, hate crimes, and bias crimes. In short, crimes that require specific intent are concerned with particular sorts of desires; crimes that require hatred of a victim are concerned with particular passions; and crimes that require bias or prejudice toward a victim are concerned with particular standing beliefs.

One who acts with a specific intention possesses a desire (however reluctantly conceived) that her actions will bring about a particular future state of affairs. Desires of this sort are motivational in the sense that they give reasons for action. In idiomatic terms, they constitute goals, or perceived goods, or ends to which actions are means. They are, in short, the objects of practical reason. To possess a specific intention is thus to have a motive for action in the sense in which that term is used in murder mysteries and detective movies.

144 Some of these distinctions were recognized in the older literature of “conceptual analysis” in philosophy. Thus, Elizabeth Anscombe distinguished “backward-looking” motives like revenge from future-oriented goals like money (Anscombe, Intention (3d edit., Oxford, 1963)), and Richard Peters distinguished motives as goals from motives as emotional states (Peters, The Concept of Motivation (Routledge, 1958)).
In contrast, to explain a defendant’s action as a product of hatred is not itself to attribute to him a desire to bring about some future state of affairs. It is, rather, to characterize his action as a product of a particular passion within which he was gripped at the time. Passions differ from the sorts of desires put at issue by specific intent crimes in that they are felt emotional states rather than ends to which actions are means. We speak of emotions “motivating” actions, saying such things as “he lashed out in anger,” or “she was in a fit of jealous rage,” or “they got lost in the excitement of the moment,” or, simply, “he lost it.” But when we say such things, we are not explaining others’ conduct by reference to their future goals. (Indeed, we are very often explaining why their powers of practical reasoning were (partially) suspended so as to make them less able to adopt appropriate means by which to achieve their goals.) Inasmuch as hatred is akin to jealousy, anger, envy, love, and the various pleasures, it motivates action in the way that passions (as opposed to desires) motivate action. It is an emotional state within which an actor acts; it is not a future state toward which an actor acts.

Let me pause in passing to make clear that there is not a clean distinction between all desires and all passions. But there is a clean distinction between the sorts of desires that are put at issue by specific intent crimes and the sorts of passions that are put at issue by hate/bias crimes. Some desires--i.e., cravings--are emotionally laden. But the desires with which specific intent crimes are concerned need not be. One is liable for the specific intent crime of burglary if in breaking and entering into another’s dwelling, one seeks to steal a television, whether or not one experiences any emotion when anticipating the achievement of that goal. Conversely, sometimes emotions provide reasons for action that enter into practical syllogisms in a manner analogous to desires, as when fear rationally gives one a reason to hide from an aggressor. But
sometimes emotions simply cause actions without the intervention of reason, as when fear causes
one to raise one’s arm in anticipation of a blow. Inasmuch as a defendant is liable for a hate
crime if his actions are caused by hatred of his victim, regardless of whether his hatred has
entered into his practical reasoning or not, hate crime liability concerns itself with the possession
of a particular passion (even when a defendant’s passion is in fact among his reasons for action).

Finally, while bias crimes and hate crimes are typically lumped together into a single
category, acts done as a result of bias or prejudice may be quite differently motivated than acts
done out of hatred. Hatred may necessarily depend upon bias, and bias may commonly cause
hatred, but bias does not necessarily depend upon hatred. One who is biased or prejudiced
against others possesses false beliefs about them. Moreover, these beliefs are what are called
“dispositional beliefs,” because they constitute standing dispositions to draw particular
conclusions about particular persons in particular situations without more particular information.
One who is biased or prejudiced against a group of persons jumps to conclusions about the
members of that group. It is because prejudices constitute sets of dispositional beliefs that we
think of them as “motivational.” They induce discriminatory actions by disposing actors to make
certain (false) factual assumptions about others within the course of reasoning about what they
ought to do vis-à-vis those others.

If this analysis of the difference between specific intentions, hatred, and bias is accurate,
then it must be admitted that hate/bias crimes punish some defendants more than others because
of certain emotions or beliefs that they harbored while acting. Before we turn to the significance

---

145 This appears true when the object of hatred is an entire group of persons defined by a morally
irrelevant trait such as race, religion, gender, etc. If persons can deserve hatred (say, by culpably
committing terrible wrongs), then it is not true when the object of hatred is a person who merits
such an emotional response.
of criminalizing these emotional and belief states, let us pause to consider two arguments that might be advanced to dispute the above analysis, and thereby, to dispute the claim that hate/bias crimes concern themselves with new and novel sorts of mens rea. First, many hate/bias crime statutes do not use the language of hate or bias at all. As such, these statutes do not appear to criminalize any emotional or belief states (whatever the merits of so doing might be). Second, it appears plausible to argue that persons who act out of hatred toward, or belief in the unsuitability of, say, Italians within their community, are seeking, by their actions, the future state of affairs in which there are no Italians in their neighborhood. Such an argument collapses emotions and prejudicial belief states into desires, and thus allows advocates of hate/bias crime liability to maintain that the mens rea of hate/bias crimes is no different than the mens rea of traditional specific intent crimes.\footnote{See Kent Greenawalt, Reflections on Justifications for Defining Crimes by the Category of the Victim, 1992/93 Annual Survey of American Law, pp. 620-25 (arguing for this collapse).}

It is not uncommon for so-called hate or bias crime legislation to avoid reference to hate or bias altogether. A number of statutes simply require fact-finders to find that the defendant chose to harm his victim “because of” the victim’s race, religion, ethnicity, etc.\footnote{For example, the proposed Federal Hate Crimes Prevention Act of 1999, while bearing reference to hate both in its title and in its statement of purpose, nevertheless failed to make a finding of hate necessary to liability. It recommended that Federal law be amended to read as follows: Sec. 4. Prohibition of Certain Acts of Violence . . . "(c)(1) Whoever, whether or not acting under color of law, willfully causes bodily injury to any person, or through the use of fire, a firearm, or an explosive device, attempts to cause bodily injury to any person, because of the actual or perceived race, color, religion, and national origin of any person— (A) shall be imprisoned not more than 10 years, or fined in accordance with this title, or both; and (B) shall be imprisoned for any term of years or for life, or fined in}
language appears neither to require a finding that a defendant hated her victim, nor to require a finding that the defendant harbored any particular beliefs about her victim as a result of the victim’s race, religion, ethnicity, etc. It simply requires a finding that the defendant used a prohibited criterion as her principle of selection. And such an inquiry appears analogous to the sort required by specific intent crimes: Did the defendant have as her purpose the victimization of someone of a particular race, religion, ethnicity, gender, etc.? It might be reasonably thought, therefore, that at least some hate/bias crimes are defined in a manner that makes them just like traditional specific intent crimes. And it might be further argued that all hate/bias crimes can and should be so defined so as to avoid any complaints that might come from inventing new mens rea categories.

accordance with this title, or both if—
(i) death results from the acts committed in violation of this paragraph; or
(ii) the acts committed in violation of this paragraph include kidnapping or an attempt to kidnap, aggravated sexual abuse or an attempt to commit aggravated sexual abuse, or an attempt to kill.

(2)(A) Whoever, whether or not acting under color of law, in any circumstance described in subparagraph (B), willfully causes bodily injury to any person or, through the use of fire, firearm, or an explosive device, attempts to cause bodily injury to any person, because of the actual or perceived religion, gender, sexual orientation, or disability of any person

(i) shall be imprisoned not more than 10 years, or fined in accordance with this title, or both; and

(ii) shall be imprisoned for any term of years or for life, or fined in accordance with this title or both, if

(I) death results from the acts committed in violation of this paragraph; or
(II) the acts committed in violation of this paragraph include kidnapping or an attempt to kidnap, aggravated sexual abuse or an attempt to commit aggravated sexual abuse, or an attempt to kill.

It would seem, however, that when hate/bias crimes are defined without reference to hate or bias, their language ill-fits both their purpose and the occasions on which they are invoked. Suppose, for example, that a defendant simply wants to steal from someone who is unlikely to put up a serious fight, and so for that reason he seeks to mug a woman. Or imagine a defendant who is looking to pick a fight and chooses an Amish person because he knows that by virtue of his pacifist commitments, his victim will not resort to physical force to defend himself. In these cases the defendants choose their victims because of their victims’ gender or religion. But these are not paradigm cases of hate crimes, nor, we hazard, would they be prosecuted as such, because they do not reflect actions done out of hatred for or prejudice against particular communities of people. Indeed, if hate/bias crimes are really committed whenever defendants simply pick their victims because of certain traits, such as gender, then all rapes are hate crimes (whatever the gender of the victim), as are all crimes that are committed by persons who use their victim’s race, religion, ethnicity, gender, or disability as proxies for other criteria (such as wealth, physical weakness, pacifism, etc.).

If prosecutors and courts do not invoke hate/bias crime statutes in cases of this sort (as they surely do not), then the practical implications of hate/bias crime legislation—even hate/bias crime legislation that does not make reference to hate or bias – is to punish defendants for hate and bias. In investigating the defensibility of hate/bias crime legislation, therefore, we are entitled to proceed by inquiring into the implications of requiring fact-finders to find not just that a defendant used a victim’s membership in a group as the principle criterion for selecting that

148 Recall the scene in the movie “Witness” in which some local boys taunt and physically assault Harrison Ford’s Amish friends precisely because they believe that the Amish will not retaliate. Witness, ____ (19__).
person, but that the defendant victimized that person out of hatred toward, or prejudice against, a
group of which she was a member. In short, we do well to ask into the results of legislation
whose language matches its point and practice, namely, punishment for acts of hate and bias. So
while the mens rea language of some hate/bias crime statutes may sound traditional, their real
mens rea requirements—hatred and bias—are far from traditional: they are emotions and
dispositional beliefs altogether different from the desires traditionally put at issue by specific
intent crimes.

The second means by which advocates of hate/bias crime legislation might analogize
hate/bias crimes to specific intent crimes is by arguing that a person who, say, acts out his hatred
toward, or bias against, women in the workplace, in fact desires and intends his actions to bring
about the state of affairs in which women are not present in the workplace.\textsuperscript{149} By collapsing
emotions and prejudicial beliefs into desires, this argument permits the conclusion that all
hate/bias crimes are exactly like specific intent crimes in criminalizing actions that are done with
further particular purposes.

There is nothing conceptually confused about this argument. Its defensibility depends
upon whether, as an empirical matter, actions that are caused by emotions or prejudicial beliefs
(with or without the intervention of practical reasoning) are in fact plausibly characterized as
actions that reflect intentions concerning future states of affairs. It seems to us that this argument
grossly over-intellectualizes the mental phenomena that the categories of desires, emotions, and
dispositional beliefs nicely delineate. While a straight defendant who manifests hatred toward a

\textsuperscript{149} See Greenawalt, \textit{supra} note 146. For a defense of the related claim that there is no morally
relevant difference between acts that satisfy external goals (e.g., to steal) and acts that satisfy
internal goals (e.g., to obtain sexual satisfaction or to express animus toward another) see
Robinson, \textit{supra} note 142, at 608.
gay neighbor, or who treats him as inferior, might, upon reflection, answer “yes” to the question, “Would you prefer a world in which there are no gays?,” this does not mean that at the time of action he entertained a conscious belief of this sort, or any desire that his actions contribute to such a state of affairs. He could simply have been manifesting a visceral emotion that had not (yet) been given propositional content, or a (dispositional) belief from which he had drawn no conclusions concerning (what are to him) optimal future states of affairs. That we armchair psychologists can identify future states of affairs with which defendants’ emotions and beliefs appear consistent is immaterial. The question is whether we have reasons to attribute intentions about such future states to persons who do not subjectively experience them. In our view, nothing is gained by so doing (except the doctrinal advantage that comes to those who may then analogize hate/bias crimes to specific intent crimes150) and much may be lost.

First, raw emotions, dispositional beliefs, and desires are experienced differently, and it obfuscates their phenomenological distinctness to collapse them into a single category. Second, these felt differences may very well be morally relevant. We may have better control over one category of mental states than over others; or we may be more (or less) culpable for harboring one category of mental states over others (regardless of our control over these states); or the law may be better able to affect our possession of one category of mental states over others (assuming, for example, that we have control over all of them, but that some are more responsive

150 For a brief defense of the claim that it would mark a “purely semantic” difference to draft hate/bias crime legislation so as to prohibit victimizing another “with the purpose of subjecting that person to racial humiliation, see Murphy, supra note 143, at 22. For a general discussion of the view that “emotions involve appraisal or evaluation of an object,” see Dan M. Kahan and Martha C. Nussbaum, Two Conceptions of Emotion in Criminal Law, 96 Colum. L. Rev. 269, 285-89 (1996). This view is often confused for the argument above; but that emotions have intelligibility conditions does not, by itself, make them goals.
to external incentives than others). For these and other reasons, we are likely to overlook morally relevant differences between actors if we translate the emotions and dispositional beliefs experienced by some into the sorts of desires that consciously motivate others.

Let us now stop to take stock of our inquiry thus far. We began with the uncontroversial observation that if hate and bias are construed as legal criteria of culpability, then they function as distinct mens rea to be proved as part of the prima facie case for hate/bias crime liability. We then established that hate and bias are unusual mens rea requirements, because they constitute “motivational,” rather than merely intentional, states of mind. To determine whether a defendant acted out of hatred or prejudice, fact-finders must ask not only whether the defendant acted purposefully, but why the defendant formed the purposes that she did. As we discussed, the only doctrinal precedent that might be invoked for this sort of motivational mens rea is specific intent—the mental state required for such crimes as burglary, attempted murder, and aggravated assault. But as we have just concluded, specific intentions, hatred, and prejudice bear markedly different phenomenological pedigrees. A specific intention constitutes a certain desired goal; hatred is a felt emotion; and bias is a set of (dispositional) false beliefs.

Advocates of hate/bias crime legislation would thus be wise to argue, not that these are identical sorts of mental states, but that laws that target these different mental states are morally identical. Inasmuch as the criminal law has long targeted motivational desires (via the enactment of specific intent crimes), and inasmuch as passions and (dispositional) beliefs are members of the same genus of motivational mental states, the morality of targeting hatred and bias (via the enactment of hate/bias crimes) might reasonably be thought to be identical to the morality of
enacting specific intent crimes. It remains for us to ask, then, whether the differences between specific intentions, passions, and (dispositional) beliefs make a moral difference.

The answer to this crucial question falls out of an appreciation of the special nature of emotions and dispositional beliefs, and specifically, the special nature of hatred and prejudice. The emotions and beliefs with which hate/bias crimes are concerned are not occurrent states of mind; they are, rather, character traits possessed by defendants over time. While one can form an intention, set a goal, and fix on a desired object in a moment's time, it is hard to conceive of what it would mean to hate or bear bias against a group only momentarily. Rather, to experience the emotion of hatred whenever one encounters members of a particular group of persons, or to regularly jump to conclusions about members of particular groups in ways that prevent careful consideration of facts unique to such persons, appears to be a matter of disposition.

Character traits are both dispositions towards certain sorts of actions and dispositions towards certain sorts of mental states. To be cowardly, for example, is to be disposed both to feel fear in (often mildly) threatening circumstances and to flee from its source. To be kind is to be disposed both to empathize with others’ conditions and to provide aid to others in circumstances in which it is not obligatory to do so. To be greedy is to be disposed both to covet things that exceed one’s share and to take things that are not one’s due. Character traits of these sorts may be a product of nature, nurture, or some combination of both of these influences. In some cases, cowardliness, kindness, and greediness appear so integral to actors’ personalities as to make one suspect that these traits are genetic. In other cases, these virtues and vices appear to be learned—the products of years of fortunate or unfortunate parenting, schooling, peer pressures, and life experiences.
We can all appreciate how someone might feel hatred toward another person (because, for example, of certain things that she has done), without being a “hateful” person—that is, without being disposed to feel hatred towards groups of persons generally, or towards the offending person in particular. We also all appreciate how someone might leap to an unjustified conclusion about another person (because, for example, one lacks the time or ability to gather further important information about him that would alter one’s first impression) without being a “biased” person—that is, without being disposed to assume that the others possess certain attributes solely by virtue of their membership in certain groups (e.g., racial, ethnic, or religious groups). Thus, hatred and bias are not necessarily, or as a general matter, dispositional.

But hatred and bias towards whole groups of persons (particularly groups defined by such morally irrelevant criteria as race, ethnicity, gender, age, religion, or disability) does appear dispositional. To experience the felt passion of hatred whenever one encounters (or even thinks about) a member of a particular group appears to reflect a standing disposition (whether learned or innate), not a momentary or occurrent emotion. And to regularly draw conclusions and reason from assumptions about the character and deeds of particular persons based solely on their membership in larger groups appears to be similarly dispositional. Thus, for example, to be homophobic is to be disposed to feel revulsion when one encounters persons who are homosexual, and to act in ways that deny them equal status and benefits within the community. To be a chauvinist is to be disposed to believe that members of a particular sex are inferior, and to act in ways that oppress them. To be a racist is to be disposed to believe that members of another race are inferior, and to act in ways that subjugate them.
The sorts of mental states (intentions concerning future states of affairs) that are targeted by specific intent crimes and the sorts of mental states (emotions and beliefs) that are targeted by hate/bias crimes thus bear significant dissimilarities. While one’s intentions are often a product of certain aspects of one’s character, one may form the intention, say, to kill without being disposed towards violence, and one may set one’s sights on another’s wealth without being a greedy person. But one cannot hate or be prejudiced against Catholics and women without being a bigot and a sexist—that is, without being disposed to feel antipathy whenever one is confronted with members of that religion and sex, or without being disposed to believe derogatory things about them and to act in ways that oppress them. In punishing a defendant for the intentions with which he acted, then, specific intent crimes do not necessarily punish a defendant for having bad character. But in punishing a defendant for hating or being prejudiced against his victim because of his victim’s membership in a particular group, hate/bias crimes do necessarily punish a defendant for having bad character. In fact, they punish a defendant solely for bad character.

Now many will be quick to say: “No they don’t. Hate/bias crimes don’t punish defendants for having bad character; only for acting on it.” But remember: the underlying criminal act, together with the mens rea to commit that act (e.g., the intentionality of the act), is already punished by existing criminal law. The enhanced penalty attached by hate/bias crime legislation is not for the underlying act, nor is it for the intentionality with which it is committed.

---

151 Indeed, according to Jeffrie Murphy, that is its point. “We look to motives not to punish them as thoughts alone but as evidence about the ultimate character of the person being punished.” Murphy, supra note 143, at 22.

152 See, e.g., Greenawalt, supra note 146, at 624 (“Actors who have the wrong kinds of feelings are not being asked to eliminate the feelings. They are asked not to commit criminal acts on the basis of their feelings.”)
it is for the hatred or prejudice that motivated the defendant to form and act on that intent. It is for the defendant’s bad character. In the next section, then, let us turn to the questions that advocates of hate/bias crime legislation must answer in order to defend the moral and political legitimacy of criminalizing character.

B. The Moral and Political Legitimacy of Punishing People for Bad Character

Most of us are confident that would-be defendants can choose not to rape, steal, and kill. We regularly blame and hold persons morally and legally responsible for intentional harms precisely because we think that persons have meaningful control both over what they intend to do and over the actions that they commit. After all, to intend to bring about a future state of affairs is simply to choose both an end and the means by which it is accomplished. And, as the word implies, a choice is precisely that: it is something over which an actor has autonomous control.

It seems much less clear to what degree people can will away, or decide not to have, particular character traits including particular emotions and (dispositional) beliefs. Can one simply decide not to be selfish, or greedy, or narcissistic? Can one simply refuse to believe something that one has long been disposed to believe? Surely not by an act of will alone. Certainly people can gather more information about others so as to make more discriminating, and thus less discriminatory judgments about them. But it is in the nature of bias for one to believe that no further information is required by which to make an informed judgment about a person. And it is unclear how one could simply will a change in one’s epistemic standards.
Persons can undoubtedly (albeit with mixed success), *indirectly* alter aspects of their character by choosing to subject themselves to experiences that promise life-changing effects. For example, many spend a great deal of money each year for counseling and therapy, with the expectation that it will help them identify and control feelings and desires that get the better of them. And persons may be able to choose to put themselves in circumstances that challenge them to behave in ways that over time affect their beliefs and desires. For example, someone might not be able to will away his chauvenism; but he might be able to rid himself of his condescension by shouldering the sorts of tasks that are commonly performed by women, so that over time his disdain for women in traditional roles fades to respect for the hardiness that it takes to repeat the multitude of mentally and physically demanding chores required to raise children and keep a home.

Perhaps those who do not seek to “build character” deserve punishment for their bad character, not because they chose their character or failed to will it to be otherwise, but because they failed to will actions that might have caused their character to be otherwise. And perhaps by punishing people particularly harshly when they do bad deeds out of hatred for, or bias against, their victims' gender, race, religion, etc., the criminal law will induce people to take actions that will indirectly alter, over time, their emotional responses to, and entrenched beliefs about, persons with such characteristics.\(^1\)

But while character is not immutable, one’s ability to affect it (either by raw choice or through what one hopes will be character-altering experiences) is clearly imperfect and

\(^1\) We do not embrace a utilitarian theory of punishment, so if this is the best justification that can be given for hate/bias crime legislation, then our objections to it will exceed those articulated in this Part of the paper.
unpredictable. Inasmuch as we cannot abandon our emotions and beliefs the way we can abandon our plans—i.e., simply by choice—criminal legislation that targets emotions and (dispositional) beliefs targets things that are not fully or readily within a defendant’s immediate control. And if the state ought not to punish us for things that we cannot autonomously control, then hate/bias crime legislation is suspect for doing just that.

Before we take up the second aspect of hate/crime legislation that is morally troubling, let us be clear about what we are not arguing here. We are not suggesting that a defendant cannot be punished for racism or sexism or homophobia because he did not choose to be a racist, or a sexist, or a homophobe. We are suggesting that a defendant cannot be punished for racism, or sexism, or homophobia because he cannot simply choose not to be a racist, or a sexist, or a homophobe. While distinguishing our argument in this way might appear initially to slice the bologna awfully thinly, it is of crucial importance. Few criminals become criminals through their own devices alone. Most turn to crime after years of tragic abuse or neglect for which parents, guardians, teachers, and perhaps the rest of us in our collective apathy are to blame. Were we to refuse to punish a defendant for character flaws solely because they were not entirely of his own making, we should refuse to punish (virtually) all defendants, because their choices were hardly entirely of their own making.

In our view, the law rightly punishes defendants who choose to do harm, even when their choices are informed by aspects of character that are not chosen, because in a very meaningful sense, they could have chosen otherwise. The law thus rightly punishes someone for his choice to throw a rock through a neighbor’s window, because at the point at which he made his choice, he could have chosen to do any number of other, non-harmful actions. But the law is much less
obviously entitled to punish someone for the hatred he bears his neighbor at the time he throws a rock through his neighbor’s window, for at the moment that he made his choice, he could not have willed away his hatred.

Notwithstanding the relative inelasticity of character, we are sympathetic to the view that moral culpability is largely a function of the reasons for which persons act and the emotions that attend their actions. Contrary to the traditional assumptions of the criminal law, it seems to us that surprisingly little is learned about a defendant’s moral culpability by discovering that the defendant intended a legally-prohibited harm or knew that he would cause it. After all, mercy killings are intentional, legally unjustified killings, but many of us consider those who are brave enough to commit them to be highly moral people—perhaps all the more moral because of their willingness to risk legal liability to end a loved one’s suffering. Inasmuch as we can distinguish the mercy killer from the contract killer only by reference to his relative motivations, and inasmuch as the mercy killer appears as non-culpable as the contract killer appears culpable, our theory of moral culpability clearly departs from our doctrines of legal culpability by weighting an actor’s motivations for action far more heavily than the intentionality of his actions.

Yet hatred and prejudice toward particular groups of people are but two culpable mental states. Are they the worst of the bunch? Are they so rotten as to justify special criminal attention? How does misogyny and homophobia compare to greed, jealousy, revenge, sadism, and cowardliness? If hate/bias crime legislation is going to remain unique in picking out particular emotions and (dispositional) beliefs as bases for increased punishment, then those who advocate its enactment must either (1) defend the claim that racial hatred or gender bias are morally worse mental states than greed, sadism, jealousy, and vengeance, or (2) advance some
reason to think that such hatred and bias are responsive to criminal sanctions in a way that greed, jealousy, sadism, and vengeance are not.

It is not clear to us that advocates of hate/bias crime legislation can sustain either of these arguments, although it also seems to us that these arguments are not as hopeless as some have insisted.\(^\text{154}\) It is true that “motives cannot be readily ranked by their degree of culpability,”\(^\text{155}\) and this confounds any easy defense of the first claim. One critic suggests that this is because all motivations to assault persons, for example, share a common denominator, i.e., the motive to humiliate, and this makes them all equally culpable.\(^\text{156}\) But this explanation is implausible: someone who renders another unconscious so as to pick his pocket is clearly not motivated by a desire to humiliate, and someone who enjoys embarrassing others is clearly not as culpable as someone who enjoys inflicting physical torture on others, even though both may be said to be motivated by a desire to humiliate.\(^\text{157}\) Another critic has suggested, conversely, that the reason

\(^{154}\) Those who are convinced that hatred and bias are no more morally reprehensible than greed, indifference, etc., include Jacobs & Potter, *supra* note 4, at 80; Alon Harel and Gideon Parchomovsky, *On Hate and Equality*, 109 Yale L. J. 507, 513 (1999).

\(^{155}\) Harel & Parchomovsky, *supra* note 154, at 513. For an interesting defense of the claim that defendants are more culpable when they target victims who are uniquely vulnerable (as is often the case in hate/bias crimes), see Kenneth W. Simons, “On Equality, Bias Crimes, and Just Deserts,” 91 J. of Crim, Law and Criminology 237, 240-43 (2000).

\(^{156}\) See Jeffrie G. Murphy, *supra* note 143.

\(^{157}\) As a general matter, the search for such common denominators is a bizarre one. See, e.g., George Fletcher, *The Place of Victims in the Theory of Retribution*, 3 Buff. Crim. L. Rev. 51, 57-58 (1999); Jean Hampton, *Forgiveness, Resentment, and Hatred*, in Jeffrie Murphy and Jean Hampton, *Forgiveness and Mercy* 43-53 (Cambridge, 1988). To say, as Jean Hampton has, that an unjustified assault is culpable because it manifests a desire to humiliate is to say one thing too many. It is like saying, as George Fletcher has, that the reason that murder is wrong is because it constitutes a means of dominating another. The reason that it is culpable to intentionally assault another (absent a moral justification) is because (unjustified) assaults are morally wrong. The reason that it is wrong to intentionally kill another (absent a moral justification) is because
that motivations cannot be clearly ranked is because they are so different as to be
incommensurable.\textsuperscript{158} We find this suggestion equally implausible, not only because the concept
of incommensurability is by itself troublesome. If all culpable motivations were
incommensurable, there would be no easy cases of comparison. But there clearly are. Someone
who is motivated to exceed the speed limit in order to get to work on time is clearly not as
culpable as the sadist who seeks to prolong the torture of a child over days.

In our view, the likely reason that motives resist cardinal classifications and even ordinal
rankings is that they are highly fact-sensitive. Just as the assessment of a defendant’s alleged
negligence turns on a detailed appreciation of the circumstances in which she acted and the
information reasonably accessible to her, so the determination of someone’s motives for action
requires one to determine the nature and the relative weight of an inevitably complex set of
motivational desires, emotions, and (dispositional) beliefs, and to assess the overall merit of their
combination. Someone who feels a sadistic satisfaction at causing strangers pain may
simultaneously be a loving family man and a committed community leader, who works hard to
cultivate a nurturing environment for his children and to better the lives of his neighbors. That
person must be compared to someone who never derives pleasure from anyone’s pain, but who is
also indifferent to his family’s and his community’s interests, selfishly weighting the importance
of his own projects above all else. The sorts of nuanced assessments that one must make to

\textsuperscript{158} Harel & Parchomovsky, \textit{sura} note 154, at 513.
determine a person’s motivational culpability, while possible, do not lend themselves to taxonomic classifications and rankings, any more than determinations of negligence lend themselves to the sorts of judicial per se rules (“Stop, look and listen”) to which Justice Oliver Wendell Holmes famously aspired. As such, it appears difficult to say that particular motivations, say racial bias or religious hatred, are per se worse than other motivations: as between some persons they probably are; as between others they probably are not.

Those who admit that hatred and bias toward particular groups may be no worse than other motivations, but who remain convinced that it is just to enact legislation that uniquely punishes such hatred and bias, must resort to the second argument provided above: namely, that the motivations of hatred and bias towards particular groups can be affected by criminal sanctions in a way that all other motivations cannot. This argument allows them to maintain that while all culpable motivations might be deserving of punishment, only some—namely, hatred and bias towards a particular race, gender, religion, etc.—can be deterred by punishment. If criminal sanctions can properly be imposed only when they will deter undesirable conduct and/or mental states (a premise to which no retributivisist would subscribe), then there is good reason to enact hate/bias crimes without enacting legislation that targets other equally culpable motivational states.

Is there any reason to think that racism, sexism, and homophobia, can be more readily purged by the criminal law than can, say, greed, sadism, or jealousy? Some maintain that

---

159 See Baltimore & O.R.R. v. Goodman, 275 U.S. 66, 70 (1927) (stating that as specific standards of conduct become clear, judges should lay them down once for all as per se rules of conduct). For Justice Benjamin Cardozo’s famous reply, see Pokora v. Wabash Ry. Co., 292 U.S. 98, 104 (1934) (arguing that what is reasonable in one set of circumstances cannot be generalized to all similar circumstances).
racism, sexism, and homophobia are learned, while sadism, vengeance, and jealousy are not. They thus argue that racism, sexism, and homophobia can be un-learned in a way that these other dispositions cannot. Perhaps this is so. Even if one believes that the genesis of group hatred and prejudice is environmental, whereas the genesis of other vices is genetic or otherwise innate, one would need reasons to believe that learned dispositions are less immutable than unlearned ones before criminalizing the former but not the latter. And one would need further proof that racists, sexists, and homophobes will more quickly, effectively, or cheaply un-learn their vices when threatened with enhanced criminal liability, than when educated or coerced in other (non-legal, or at least non-criminal) ways.

If advocates of hate/bias crime legislation are unable to defend either the claim that hate and bias are more culpable than other motivations or the thesis that hate and bias are uniquely susceptible to coercion by criminal penalties, then they have but two avenues of argument available to them. They must either (1) admit and defend the fact that hate/bias crime legislation arbitrarily picks out for extra punishment a set of mental states that are a subset of a larger class of equally vicious states, or (2) convince theorists and legislators to generalize hate/bias crime legislation by radically revising our culpability doctrines so as to take into account, and dish out punishment in proportion to, all culpable emotional and (dispositional) belief states that motivate defendants to do criminal deeds—from racial bias to jealousy to road rage to cowardice.

The first line of argument is seemingly foreclosed, for it implies that the criminal law both does not and need not treat like cases alike. Equally culpable defendants who commit equally wrongful acts (for example, one who throws a rock as a result of vengeance and one who throws a rock as a result of racial hatred) can justifiably be punished unequally. While like cases
must be judged alike, on pain of irrationality, the argument here is that like cases need not be punished alike. Those who would make such a claim clearly cannot think that equality of treatment is much of a moral value. They must believe that so long as someone is not punished more than he deserves, he cannot complain that he is punished more than someone of equal culpability and wrongdoing.

There are at least two powerful responses to such a position. The first asserts that equality of treatment is a significant moral value, and that instances in which persons are punished unequally for committing equally wrongful deeds with equal culpability are instances of serious injustice. We will not here rehearse the arguments for the moral worth of equal treatment that reside in the vast literature on equality and comparative proportionality. Suffice it to say that one who seeks to defend the unique status of hate/bias crimes within our criminal law may well be forced to answer the challenges within this larger literature.

The second response is to agree that no injustice is done to a defendant who receives his due, even when he is punished more than another defendant who committed an identical wrong with equal culpability. The injustice, rather, resides in the fact that the other is punished less than his due. While no defendant can complain when he gets what he deserves, and while few would complain when they get less than they deserve, it is a violation of retributive justice to

---

allow some to be punished less than they deserve. On this argument, if persons who commit crimes as a result of racist, sexist, or homophobic motivations deserve more punishment than they will get under traditional criminal statutes, then it is not unjust to enact legislation that allows adjudicators to give them the punishments that they deserve. But it is unjust not to enact simultaneously legislation that allows adjudicators to give other, equally culpable persons their due. For if persons who commit crimes out of hate deserve enhanced punishments, and if persons who commit the same crimes out of vengeance are admittedly as culpable as those who commit them out of hate, then persons who commit such crimes out of vengeance deserve the same enhanced penalties. The failure to punish them equally thus constitutes a violation of retributive justice. On this argument, equality is not a value, but retribution is. If hate/crime statutes are justified, and if other motivational states are as culpable as hate and bias, then hate/crime statutes cannot be justified alone. Rather, their enactment reveals a larger flaw in our system of penalties; namely, that current penalties do not give highly culpable defendants their due. We should learn the lesson that hate/bias crimes have taught us—namely, that many ill-motivated defendants are under-punished—and enact more general legislation that allows adjudicators to accomplish retributive justice by punishing all equally culpable wrongdoers equally.

This is effectively the final argument that advocates of hate/bias crimes might make in response to the charge that such crimes cannot justifiably continue to enjoy a special status within the criminal law. If one believes that an actor’s moral culpability is better measured by his background motivations than by the intentions, knowledge, or degree of conscious awareness that he possessed concerning the results of his conduct, then one might well think that our
criminal law should be radically revised so as to better mirror the conditions of true moral culpability. Fact-finders should be required generally to assess and punish in proportion to the desires, emotions, and (dispositional) beliefs that prompt defendants to do wrong. Hate/bias crime legislation is justified, on this argument, as an interim experiment that successfully demonstrates how legislation can and should more generally permit adjudicators to punish defendants in proportion to the relative culpability of their motivational mental states.

Such a suggestion brings us to the third moral problem that confronts both those who defend hate/bias crime legislation in particular, and those who are tempted by the proposal that we revise mens rea requirements generally so as to punish defendants for wrongdoing committed with culpable desires, emotions, and (dispositional) belief states. Inasmuch as hate/bias crime legislation in particular, and motivationally-oriented legislation in general, ultimately punishes persons for standing traits of character, it is best explained and justified by what is often called a “character theory of the criminal law”—a theory that takes the proper goals of the criminal law to be the punishment of vice and/or the cultivation of virtue.\(^{161}\) Inasmuch as these are distinctly illiberal goals, those who advocate hate/bias crime legislation, or who are persuaded to generalize the sort of inquiry that it permits, cannot simultaneously count themselves political liberals.

Political liberals classically distinguish a theory of the Right from a theory of the Good, and confine the state to legislating in accordance with a theory of the Right. They maintain that

\[\text{\footnotesize\textsuperscript{161} For a general defense of the role of character assessments in justified impositions of criminal liability, see Peter Arenella, Character, Choice and Moral Agency: The Relevance of Character to Our Moral Culpability Judgments, 7 Soc’l Phil and Policy 59 (1990), reprinted in Crime, Culpability and Remedy., E.F. Paul et. al. eds (Cambridge, 1990). For a general critique of the character view of responsibility, see Michael Moore, Choice, Character, and Excuse, 7 Soc’l Phil. and Policy 29 (1990), reprinted in Moore, Placing Blame, supra note 5.}\]
the state may properly use its power only to construct and protect a fair framework of cooperation, defined as a system of rights that allows all citizens maximal equal liberty to pursue their own unique conceptions of the good life. Common to contemporary liberalism is the claim that reasonable people can, in principle, agree to disagree, and can, in practice, agree to peaceable institutional means by which to disagree; but they cannot agree. That is, they can agree on, and thus give consent to, state action that preserves their ability to pursue their own unique conceptions of a good life, but they cannot, and will not, agree on any singular conception of the good life, and thus, they cannot, and will not consent to state action that advances any particular conception of the good life. Inasmuch as state action is both theoretically unjust and practically impotent unless it can command the consent (if only hypothetically) of reasonable persons, state action is illegitimate if it pursues a particular conception of the good life.\textsuperscript{162}

Liberals have long believed that theories concerning virtuous and vicious character traits (and how to cultivate or eliminate them) belong to the province of the Good, rather than that of the Right. First, it appears that one person’s vice is another’s virtue. That is, the vices and virtues of character are not among the things about which people agree, whatever else they believe. Reasonable people can seemingly disagree over whether persons should worship God, or honor blood ties, or adopt a vigorous work ethic, or be so honorable as to keep very burdensome promises, or elevate truth or others’ emotional well-being.

Second, as we have already discussed, character traits appear only indirectly responsive to choice. Inasmuch as punishment for character traits smacks of strict liability, strict liability threatens to chill liberty, liberals typically predicate moral responsibility on matters of choice. Because persons cannot choose their character traits, but can affect them only indirectly and imperfectly, vices and virtues have appeared to liberals to be inappropriate candidates for praise and blame.

Third, persons can seemingly agree to disagree without settling the question of the sorts of character traits that persons ought to cultivate or suppress. Indeed, until the enactment of hate/bias crimes, our criminal law was testament to the fact that we could meaningfully assess actors’ culpability without resolving inevitable disputes over the sorts of traits that people should cultivate. While people no doubt disagree about the relative culpability of acts of spite, jealousy, vengeance, and racism, few have disagreed that purposeful harms are more culpable than harms committed only knowingly; and that harms committed knowingly are more culpable than harms committed recklessly or negligently. Inasmuch as contemporary liberalism is wedded to the view that our peaceable co-existence requires the state to limit its scope to matters upon which we can agree, it is committed to the view that our old ways are better than our new ones—that hate/bias crime legislation exceeds the proper bounds of state action, as would any attempt to radically revise mens rea doctrines so as to use state power to affect people’s moral character.

Those who persist in thinking it appropriate for the state to criminalize certain aspects of character must admit, then, that they are not liberals. Instead, they are probably political perfectionists, who, broadly speaking, consider it appropriate for the state to use its power to
perfect its citizens morally.\textsuperscript{163} Perfectionists consider it legitimate for the state to enact legislation that will induce us to become more virtuous and less vicious by, say, nurturing in us charitable, kind, and courageous dispositions, and suppressing selfish, cowardly, cruel, or racist dispositions.\textsuperscript{164} That reasonable persons may disagree about the list of virtues to be cultivated and vices to be suppressed is not, by itself, a reason for the state to stay its hand. Indeed, as perfectionists maintain, the state cannot, and should not, remain agnostic between competing conceptions of the good. Legislators both inevitably will, and morally must, appeal to particular conceptions of the good when legislating conduct. The political morality of their choices thus turns on their success in deciphering what is, in fact, good for persons.

The willingness to use the criminal law to punish people for bad character and to encourage people to cultivate good character is at least closely compatible with, if not required by, the perfectionist directive to use the power of the state to affect individual virtue when possible. Those who advocate the enactment of legislation that punishes racism, sexism, homophobia, and so forth, would thus be wise to align themselves politically with those who

\textsuperscript{163} For recent constructions of perfectionist political theories that defend the extension of state action to the cultivation of virtue while still promising “liberal-like” results, see Thomas Hurka, \textit{Perfectionism} (Oxford University Press, 1993); Steven Wall, \textit{Liberalism, Perfectionism, and Restraint} (Cambridge University Press, 1998).

\textsuperscript{164} We take communitarianism to be a species of perfectionism, for it assigns to the state the task of perfecting citizens qua citizens. Communitarians differ, on this construction, from Aristotelian perfectionists in that they do not license the state to perfect citizens qua moral persons. That is, they accord the state the power to cultivate in its citizens the other-oriented virtues that citizens should have in order to be productive, contributing members of a political association, but they do not grant the state the power to cultivate in its citizens the self-oriented virtues that citizens should have to be fully moral persons. Aristotelian perfectionists, on the other hand, consider the entire range of virtues and vices to be the legitimate jurisdiction of the state, curtailing state power only when law turns out to be a clumsy or otherwise ineffective means of affecting virtue.
view the inculcation of virtue and the elimination of vice as a legitimate state goal. However, even those who are, in principle, willing to criminalize certain aspects of character must honor what Joseph Raz calls “the morality of freedom.” As a political perfectionist, Raz has powerfully insisted that in many instances what is good (virtuous) for persons is a product of what they autonomously chose to do and to believe. Hence, a legislator who seeks to pursue a perfectionist agenda must be careful to preserve a significant arena in which persons may make their own choices about what is good. Inasmuch as many persons come to make good choices only by making a certain number of bad ones, the state must tolerate a certain number of bad choices in order to maximize the moral virtue that derives from exercises of liberty.  

Not only must those who are willing to criminalize hate and bias meet the challenges of contemporary liberalism generally, but they must be confident that in enacting such legislation, they are not inhibiting a liberty to be bad that is necessary to the cultivation of good. Our long reluctance to racist, sexist, homophobic, and other intolerant associations suggests that such associations, while not themselves of moral worth, may be among those unfortunate things that we must tolerate in order to maximize pursuits that are of moral worth. They may contribute to an important marketplace of ideas (if only by revealing how ugly some ideas can be), or they may be such that we must tolerate them on pain of chilling other associations that are of moral worth, or the precedent set by prohibiting them may motivate legislators to prohibit other associations that are in fact quite worthy. Whatever explains our historical commitment to tolerating the non-violent activities of pornographers, anti-Semites, the Ku Klux Klan, the Neo-Nazis, and other sexist, racist, homophobic, and religiously intolerant groups, may be a good

reason for perfectionists to conclude that the vices of group hatred and prejudice are not best repressed by criminalizing them.

IV. THE EQUALITY ANALYSIS

According to Alon Harel and Gideon Parchomovsky, defendants who commit hate/bias crimes are neither greater wrongdoers nor are they more culpable than otherwise-motivated defendants. Harel and Parchomovsky argue that if these facts suggest that hate/bias crime legislation is indefensible, then this is so much the worse for defenses that confine themselves to considerations of wrongdoing and culpability. In their view, hate/bias crime legislation reveals the shortcomings of the long-held view that defendants may be punished differently if and only if the wrongs they commit are of different sorts or the culpability with which they commit identical wrongs are of different degrees. Instead, Harel and Parchomovsky argue for the adoption of “a fair protection paradigm” that permits (indeed, requires) the state to use criminal penalties to equalize (some) individuals’ vulnerability to crime. They describe this new paradigm as follows:

The fair protection paradigm is predicated on the principle that the criminal law is a principal means by which society provides protection against crime to potential victims. On this view, protection against crime is a good produced by the criminal justice system which, like all other state-produced goods, should be distributed in an egalitarian manner.

166 See Harel & Parchomovsky, supra note 154, at 509, 518-23.
Accordingly, the fair protection paradigm requires the state to take into account disparities among individuals in vulnerability to crime when determining their entitlement to protection. Thus, under the fair protection paradigm, victims who are particularly vulnerable to crime may have a legitimate claim on fairness grounds for greater protection against crime. Bias crime legislation, on this view, is aimed at protecting individuals who are particularly vulnerable to crime because of prevailing prejudices against them.\footnote{Id. at 509 (footnote omitted).}

According to Harel and Parchomovsky, a person’s vulnerability to crime can be calculated by multiplying the probability that she will be the victim of criminal wrongdoing by the magnitude of that wrongdoing. Some persons are more vulnerable to crime for reasons of their own choosing; others because of plain bad luck; and still others because of factors that no government could reasonably affect. In Harel’s and Parchomovsky’s view, the state cannot and should not be expected to eliminate all sources of vulnerability to crime, particularly those that stem from exercises of choice or sheer luck.\footnote{For a indictment of the assumption implicit in Harel’s and Parchomovsky’s thesis that persons have obligations to take precautions against others wrongdoing, on pain of properly bearing the burden of that wrongdoing, see Heidi M. Hurd, \textit{Is It Wrong to Do Right When Others Do Wrong?} 7 Legal Theory 307 (2001).}

But the state can and should “redress disparities in vulnerability to crime that result from certain immutable personal characteristics of the victim.”\footnote{Hared and Parchomovsky, \textit{supra} note 154, at 510.} Since some persons within our society are made more vulnerable to crime than others
because of their race, religion, ethnicity, sexual preference, gender, etc., legislation designed to increase their protection is both necessary and legitimate.

Harel and Parchomovsky argue that there are two ways in which legislators can increase the protections that are owed to particularly vulnerable groups. First, they may impose harsher penalties on those who commit crimes against members of such groups. Second, they may devote more resources to detecting and prosecuting defendants who target such victims.\textsuperscript{170} According to Harel and Parchomovsky, the latter strategy is an “impossible” one when the vulnerability to crime is attributable to immutable characteristics.

It would be futile on the part of the police to invest resources in detecting perpetrators of hate crimes because it is only after the perpetrator of the crime is detected that one can clearly discern her motives. Thus, in the context of hate crime it is easier and more efficient to use differential sanctions rather than differential enforcement efforts to reduce disparity in victims’ vulnerability.\textsuperscript{171}

The defense of hate/bias crime legislation advanced by Harel and Parchomovsky is, at first glance, both creative and philosophically tempting. There is little doubt that some communities bear more than an equal share of the nation’s crime, and that they do so solely because their members possess certain characteristics that are both immutable and morally irrelevant. It indeed seems unfair, for example, that homosexuals are at risk for being victimized

\textsuperscript{170} \textit{Id.} at 510.

\textsuperscript{171} \textit{Id.} at 510, 537.
in all the ways that heterosexuals are at risk, and then are also at risk for being victimized because they are homosexual.\textsuperscript{172}

But there are a set of reasons—both moral and conceptual—that ultimately require us to reject the “fair protection paradigm,” and with it Harel’s and Parchomovsky’s claim that hate/bias crime legislation is justified because it helps to equalize the distribution of crime.\textsuperscript{173}

First, it is Machiavellian to think that there is no moral difference between reducing crime by adding to the police force and reducing crime by increasing the penalties of those who are caught. By arguing that enhanced penalties can be imposed on defendants as a means of deterring others from perpetrating crimes on vulnerable victims, Harel and Parchomovsky are licensing the state to make examples of some defendants by meting out more punishment to them than they deserve. But why stop (or start) with those who are guilty of hate/bias crimes? We can surely reduce crimes of all sort, and against all sorts of victims, by inflicting Draconian punishments on those offenders who are unfortunate enough to be caught.

Put in philosophical terms, the first problem that besets the theory of hate/bias crime legislation that Harel and Parchomovsky advance is that, in its unabashed consequentialism, it proves too much. By seemingly assuming that there are no side-constraints to the means that a state might adopt in order to achieve the consequence of equalizing victims’ vulnerability to

\textsuperscript{172} But see Susan Gellman, \textit{Sticks and Stones Can Put You in Jail, But Can Words Increase Your Sentence? Constitutional and Policy Dilemmas of Ethnic Intimidation Laws}, 39 UCLA L. Rev. \textsuperscript{___}, 385-88 (1991) (arguing that laws designed to protect vulnerable communities, while benevolently motivated, carry implicit patronizing and paternalistic messages that add to, rather than alleviate, the burdens of those who are vulnerable).

\textsuperscript{173} For a persuasive defense of the further claim that Harel’s and Parchomovsky’s egalitarianism, however morally and conceptually compelling, is not in harmony with, and thus cannot account for, the doctrinal details of hate/bias crime legislation, see Simons, \textit{supra} note 155, at 240-43.
crime (such as inflicting undeserved punishment), Harel and Parchomovsky invite all of the
counterexamples that can be brought to bear against any purely consequentialist theory of
punishment.\footnote{See supra text accompanying notes 118-120, 124-125. For a compelling retributivist critique
of Harel’s and Parchomovsky’s consequentialist argument, and for an interesting defense of the
claim that retributivists could accommodate the egalitarian concerns behind the Fair Protection
Paradigm, see Simons, id. at 243-65.} They appear committed to punishing the innocent, letting the guilty go free,
carving up the healthy to save the sick, and so forth, so long as the consequences of so doing are
good (in this case, equality-producing), all consequences considered. But as the familiar
counterexamples make clear, it is not a matter of moral indifference whether legislators add
police officers to a beat or cut off thieves’ hands, even if the latter practice accomplishes more
crime prevention than any number of police officers could. And it is not a matter of moral
indifference whether legislators choose to commit greater resources to law enforcement or to
enhance the penalties of some offenders beyond what they deserve, even if the latter practice
accomplishes far more (hate/bias) crime prevention than any number of police officers could.

Inasmuch as they appear prepared to abandon desert as a necessary condition of
punishment, Harel and Parchomovsky cannot market their theory to those who embrace either
retributivist or mixed theories of punishment. And it is not open to Harel and Parchomovsky to
argue that hate/bias crime defendants \textit{deserve} greater punishment when their punishment is an
efficacious means of equalizing others’ vulnerability to crime. If culpability and wrongdoing
remain the only two ingredients of moral desert, and if Harel and Parchomovsky are right in
insisting that hate/bias crime offenders are neither more culpable than other defendants, nor
greater wrongdoers, then the greater punishments imposed on such offenders cannot be said to be
deserved. And Harel and Parchomovsky cannot maintain that theirs is a new theory of \textit{desert},
without abandoning any morally meaningful notion of desert. After all, they would then have to argue that if a defendant’s acquittal for a heinous offense for one reason or another reduced a community’s vulnerability to hateful or prejudicial treatment (as O.J. Simpson’s acquittal arguably reduced the susceptibility of African-Americans to hateful or prejudicial police conduct), such a person deserves no punishment for the terrible harm he so culpably perpetrated. And they would seemingly have to argue that if the state could effectively reduce the disparate victimization of a community by executing hate-motivated vandals, such vandals deserve to be killed. Since these claims are patently absurd, Harel and Parchomovsky would do better to argue that desert is not a necessary condition of punishment than to argue that persons deserve whatever punishment efficaciously eliminates the unequal distribution of crime.

One might think, then, that the only audience that will be receptive to the theory advanced by Harel and Parchomovsky is the relatively small group of criminal law theorists who embrace a pure utilitarian theory of punishment. But this philosophical congregation, while comprised of consequentialists who are prepared to sacrifice what is good in order to maximize it, should be similarly unsympathetic to their theory. For utilitarians conceive of utility as the good to be maximized, while Harel and Parchomovsky conceive of equality in the distribution of crime as the good to be maximized. Inasmuch as utility and the equal distribution of crime are fundamentally different goods, a theory that seeks to maximize the former may be in conflict with a theory that seeks to maximize the latter.

Consider, for example, two proposed penalty systems. One system will punish all offenders harshly, and will thereby reduce crime dramatically but unequally. That is, while crime rates will plummet, some communities will continue to bear more than an equal share of
the crime that persists. The second system will punish only hate/bias crime offenders harshly, and will thereby eliminate the disparate impact of crime, but will not affect an overall reduction of crime that is anything like that accomplished by the first system. Utilitarians who count each for one and only one, and who thus sum utility across the entire community, will plausibly conclude that greater net utility will derive from dramatically reducing the over-all crime rate than from equalizing it at a higher level. They will thus advocate the adoption of the former system. And they will certainly do so if the latter is more expensive to administer than the former (which it plausibly would be, since fact-finders in the latter system would have to engage in motivational inquiries in a way that fact-finders in the former system would not).

Harel and Parchomovsky, on the other hand, have to prefer the latter system to the former one, even if it costs considerably more to administer. They have to sustain the claim that a community in which men and women are frequently, but equally brutalized is better than a community in which very few are brutalized, but the few who are so victimized are all women. They have to think that if a government confronts a choice between reducing crime and redistributing the results of crimes so that males, whites, Protestants, and heterosexuals are more often victimized (relative to today’s base rates) while women, blacks, Jews, and homosexuals are proportionately less often victimized, the latter choice is the just choice. And, again, they have to reach this conclusion even if we suppose that it costs three times more for a government to equalize crime than to reduce it.

Inasmuch as utilitarians seek to maximize utility, rather than equalize it, they have to find these claims antithetical to their theory of justice. Harel and Parchomovsky thus cannot claim to
ally themselves with utilitarian theorists, nor can they hope to motivate such theorists to adopt their fair protection paradigm.

In sum, Harel and Parchomovsky’s consequentialism puts them at odds with retributivists and mixed theorists, who are both sufficiently deontological in their thinking to deny that social goods (like fair protection) can be purchased with the infliction of undeserved punishment; and their egalitarianism puts them at odds with utilitarians, who take maximizing social utility (summed across all members of society) to be a far greater good than equalizing the distribution of wrongdoing. Since most egalitarians are not consequentialists and most consequentialists are not egalitarians, few theorists can accept the theory of hate/bias crime legislation propounded by Harel and Parchomovsky without abandoning their more fundamental moral commitments.

Harel’s and Parchomovsky’s proposal thus requires a punishment scheme severely at odds with all three of the dominant theories of punishment (retributivist, mixed, and utilitarian). They might say, in their defense, so much the worse for such theories of punishment that they cannot be squared with our egalitarian ideal. So let us ask after the supposed goodness of that ideal on its own terms, leaving aside its implications for both desert and utility.

Equality in the distribution of goods and bads (benefits and burdens) is a plausible goal for legislators in a liberal state (although there are those who would famously disagree even with this ). But we should ask more particularly after the good that Harel and Parchomovsky seek to distribute equally. They tell us that the good to which each citizen is equally entitled is protection from crime. The state produces and distributes such a good when it institutes crime-prevention measures, including punishment.

See supra note 160.
On one interpretation of Harel’s and Parchomovsky’s egalitarian ideal, each person is entitled to have the state make the same efforts to prevent her victimization as are made by the state to prevent anyone else’s victimization. However, as Harel and Parchomovsky remind us, we are not all equally vulnerable to crime. So presumably their ideal is better cast in terms of a right by each of us that the state take crime-prevention measures in direct proportion to our individual vulnerabilities to crime. If we are poor, Hispanic, gay, old, female, or disabled, we are more vulnerable to crime than are many others, and thus, we deserve greater crime-prevention efforts by the state to counteract our vulnerability.

There are serious difficulties lurking in the implicit claim that we can and should measure the relationship between a set of penalties imposed on wrongdoers and a reduction of vulnerability to crime on the part of potential victims. We have a good sense of how the state can benefit individuals by allocating wealth or services to them. We thus have a good sense of how persons could demand from the state greater police protection or more resources by which to fund community watch programs. But it is very hard to get a sense of how the state can benefit one individual by punishing another. How do we assess the protection provided to future possible victims by the enhanced punishment of a homophobic assailant? Must that punishment better protect: the victim of that crime? All homosexuals who live in or frequent the area in which the assault took place? All homosexuals? All of us? And even when we fix the person or community of persons who is to be benefited by an offender’s enhanced punishment, it remains difficult to imagine how we could measure the success with which the offender’s increased penalty reduces that community’s vulnerability to crime.
There are also serious difficulties lurking in the ability to make sense of the other side of the proportionality relation sought by Harel’s and Parchomovsky’s egalitarian ideal: Although gays as a class may suffer a greater risk of crimes of violence compared to the base rate for the population as a whole, the differential vulnerabilities within that class are much greater. While being gay may be one characteristic that increases one’s chance of victimization, there are innumerable other characteristics that affect one’s likely victimization: being rich, being poor, being short, being fat, wearing glasses, carrying a wallet, being coincidentally at the scene of a crime, and so forth. If the ideal of distributive justice requires the state to equalize persons vulnerabilities to crime (as much as is feasible), then the state must both seek to measure, and have a means of measuring, each individual’s vulnerability to crime, all relevant characteristics considered.

But let us bypass these thorny issues for a more fundamental problem: What is the good in proportioning state crime-prevention efforts so as to compensate for differences in individuals’ vulnerabilities to crime? The most obvious answer for Harel and Parchomovsky is that the good is in equalizing the risk of becoming a victim of crime. On this view, we are each entitled to the same risk of crime, no more and no less, as that suffered by everyone else. It is bad to be subject to a risk of crime, good to be free of such a risk, and thus just for all of us to have an equal distribution of these bads and goods.

The problem for this most obvious answer is that it is literally incoherent. Such a view presupposes that being risked is itself something bad, and that freedom from this bad state of affairs is therefore intrinsically good. But we must ask: What is wrong (bad, unfortunate, burdensome) about being risked with violent crime? If the crime never happens, and one never
knows of one’s danger, is one hurt in any way by being “hit” with a risk? Sticks and stones can
break your bones, but risks (by themselves) can never hurt you. 176

So it cannot be that it is intrinsically good not to be risked. Rather, it is, at best, only
instrumentally good not to be risked, where the intrinsic good is not to be harmed by criminal
wrongdoing. Risks are only heuristic (epistemic) means by which we predict the occurrence of
real bads—that is, incidents of crime.

A second and better construal of the good to be distributed by state crime-prevention
efforts is thus freedom from criminal wrongdoing. Risk assessments constitute the best means
possible for legislators to ascertain how to achieve an egalitarian distribution of this good. And
since, when it comes to heuristic guides, nothing succeeds like success, heuristic aids are
wherever legislators find them. If being a woman, or a gay man, or a Catholic, allows legislators
to predict that one will be victimized more often than others, then this is relevant to how they
should manage crime-prevention efforts—but no more so than is being over-weight, poor, old, or
short, if these characteristics also allow legislators to predict disproportionate victimization. The
good is for no one person to suffer more crime than anyone else, and any indicator that someone
might be forced to do so should be considered by legislators when designing and allocating
crime-prevention measures.

This construal of Harel’s and Parchomovsky’s ideal makes conceptual sense, but does it
make moral sense? Consider two variations. On the first, suppose we have a population of 1000

176 Those who find this claim mysterious will not find the hasty advancement of it here
adequate. For detailed defenses of the thesis that risks, themselves, are not wrongs, see Hurd,
The Deontology of Negligence, supra note 18; Hurd, What in the World is Wrong?, supra note
16, at 196-201. See also Stephen R. Perry, Risk, Harm, and Responsibility, in Philosophical
Foundations of Tort Law 321 (David Owen, ed., 1995) (arguing, similarly, that risks are not
harms).
persons, 800 of whom are heterosexual and 200 of whom are homosexual. There are 100 violent
assaults each year on 100 different victims. Legislators have available to them two different
enforcement schemes that yield two different distributions of the 100 assaults: The first scheme
results in 50 assaults on gay victims and 50 assaults on heterosexual victims; the second scheme
results in 20 assaults on gay victims and 80 assaults on heterosexual victims. Is it at all plausible
to prefer the second scheme to the first? Why does not each person, gay or straight, have an
equal claim to not being a victim of such an assault? Remember, the good is not freedom from
risk (as a member of some group). The good is rather freedom from assault. Whether gay or
straight, one hundred persons are denied that good under both enforcement schemes, and on the
assumption that all persons are morally equal, we should be indifferent between the schemes.

Let us now consider the second variation. Again imagine a population of 1000 persons
with a fixed number of 100 annual assaults regardless of which of two enforcement schemes is
adopted. On this variation, however, the first enforcement scheme causes 100 individuals to
each suffer one assault, while the second enforcement scheme causes only 20 individuals to each
suffer five assaults. It seems highly plausible to argue that these 20 individuals have a legitimate
complaint against the state if it opts for the second enforcement scheme over the first, for these
individuals appear to bear more than their share of society’s crime.

Having now isolated a coherent, morally plausible ideal of distributive justice here, the
question for Harel and Parchomovsky is whether their proposal does not look more like the first
variation than the second. Hate/bias crime legislation is not like repeat-victim legislation, for it
is not at all tailored to prevent the same individuals from being repeat victims of crime. Rather,
if, ex hypothesi, it does not punish defendants for unique sorts of wrongdoing or culpability, then
it appears simply to protect *groups* from disproportionate *risks* of crime. And if this is its purpose, then it is either conceptually incoherent (in concerning itself with the distribution of risks, rather than harms) or morally indefensible (in using group membership, rather than individual victimization as the trigger for state protection).

One might sensibly seek to escape these difficulties by arguing that risks, when perceived by their victims, are harms in themselves (as the law of assaults has long recognized). One might further argue that when offenders perceive and capitalize on victims’ vulnerabilities (risks) they are more culpable for doing wrong to those victims than they would otherwise be (whether their victims perceive their own vulnerabilities or not). Thus, increased risk, *when perceived by victims or offenders*, properly increases the punishment owed to those who realize that risk.

These are plausible claims. But they are not claims about distributive justice, and they are not claims that Harel and Parchomovksy can make to rescue their theory. These are claims about substantive proportionality. They reflect the normative thesis that perceptions of risk can increase a defendant’s *desert* in two ways: if a defendant commits a crime that causes persons to perceive (and fear) future victimization, then that crime is more wrongful than it would otherwise be; and if a defendant chooses his victim as a result of perceiving her to be uniquely vulnerable (i.e., high-risk), then that defendant is more culpable for his wrongdoing than he would otherwise be. Harel and Parchomovksy cannot make either of these claims without giving up both their negative thesis (that hate/bias crime legislation cannot be explained as a proportionate response to greater wrongdoing and/or greater culpability by the perpetrator) and their positive thesis (that hate/bias crime legislation can be successfully defended as a means of doing distributive justice). And since our discussion of the theory advanced by Harel and
Parchomovsky reveals that at least one of these claims must be true in order for hate/bias crime legislation to be conceptually coherent and morally legitimate, their inability to make either claim renders their theory indefensible.

CONCLUSION

We have worked through, in considerable detail, the different justifications that can be teased out of the past decade’s worth of scholarship on the groundbreaking enactment of hate/bias crime legislation. We have canvassed claims that crimes committed out of hatred and prejudice do more harm (or harms of special moral concern) to principal victims, to the communities of which those victims are members, and to our larger society. We have explored arguments that the enactment of hate/bias crime legislation constitutes a crucial means of “sending a message” that contradicts the pernicious messages implicit in crimes motivated by racism, sexism, homophobia, and the like. We have worked through reasons to think that hatred and bias are uniquely culpable mental states that make those who are motivated by them more blameworthy than other defendants for the crimes that they commit. And we have probed the claim that hate/bias crime legislation is an important means of equalizing the distribution of the risk of becoming a crime victim by protecting communities of people who are uniquely vulnerable to crime. As we have demonstrated, each of these claims rest on premises that are empirically, morally, or conceptually suspect. As such, none, as yet, provide the necessary theoretical legitimacy for this politically popular form of criminal legislation. While little could be more important than breaking down the barriers of racism, chauvinism, bigotry, and
homophobia, it is incumbent upon the state to restrict its means to those that can be justified by our best theory of the criminal law. As this article has reluctantly suggested, the case remains to be made that hate/bias crime legislation can be so justified.