DAN M. KAHAN*

TWO LIBERAL FALLACIES IN THE HATE CRIMES DEBATE

(Accepted 26 November 2000)

Those for and against hate crime laws debate each other in the shadow of John Stuart Mill. Both accept the Millian premise that the state is justified in coercing an individual only to prevent harm to others and not to condemn that individual for holding objectionable beliefs, values, or preferences.[1]

They purport to disagree only about what that principle – the guiding tenet of modern liberalism – entails. Because hate crime laws distinguish between otherwise identical assaults based solely on offenders' motives, the opponents of those laws see them as singling out hate criminals for additional punishment solely because of their noxious ideologies. The proponents of hate crime laws, in contrast, disclaim any interest in punishing offenders for their values and insist that such laws are warranted strictly by the greater harms that hate crimes inflict, both on their victims and on third parties. At that point, the debate usually turns empirical. Is the "greater harm"

* Professor, Yale Law School. I owe thanks to Amy Baehr for her insightful response paper; to Bruce Ackerman, Jack Balkin, Kenworthey Bilz, Owen Fiss, Bernard Harcourt, Peter David Strauss, and Richard Pildes for comments; and to Anjan Sahni for research assistance.

[1] See John Stuart Mill, "On Liberty," in J.M. Robson (ed.), *Collected Works: Essays on Politics and Society*, vol. 18 (Toronto: University of Toronto Press, 1977), pp. 213, 223–224. The contrast between harm and values is a gloss on the distinction Mill draws between other-regarding acts and self-regarding actions. It is more systematically developed by contemporary theorists, who stress the inconsistency between liberalism and the condemnation of aberrant values and preferences. See, e.g., Ronald Dworkin, *A Matter of Principle* (Cambridge: Harvard University Press, 1985) pp. 191, 203; Raz, *The Morality of Freedom* (Oxford: Clarendon Press 1986), p. 420; Joel Feinberg, *The Moral Limits of the Criminal Law: Harmless Wrong-Doing* 4 (1988), p. 154; Elena Kagan, "Pivate Speech, Public Purpose: the Role of Governmental Motive in First Amendment Doctrine," *University of Chicago Law Review* 63 (1986), p. 512.

Law and Philosophy **20:** 175–193, 2001.

argument factually sound? Do the political sponsors of hate crime laws "really" believe that argument, or is the "greater harms" claim just a pretext for their illiberal desire to use state power to condemn values they abhor?

My goal in this essay is to extricate the hate crimes debate from this pattern of argument. The moral legitimacy of hate crime laws, I'll try to show, doesn't depend on whether those laws punish "harms" or instead punish "values." Indeed, it can't possibly depend on that. For unless we take an actor's motivating values into account, we lack the normative resources necessary to identify and evaluate the harm imposed by her actions. By disparaging hate crime laws for condemning offenders' motivating values, the critics subject such laws to a test incompatible with the basic theory of culpability that animates the criminal law. By suggesting that hate crime laws pass this test, the proponents of theses law advance a conceptually insupportable claim, the weakness of which does unnecessary discredit to the provisions they are defending.

The only way to figure out whether hate crime laws are morally legitimate is to determine whether the assessment they make of offenders' motives is right or wrong. Do hate crimes in fact express valuations that are more reprehensible than those expressed by other types of violent crimes? That's the question the proponents and opponents of hate crimes laws ought to be debating. The short shrift that question has received in the hate crimes debate is symptomatic of the malignant influence of liberalism – or at least one conspicuous version of it – on the formation and evolution of the criminal law.

My argument will unfold in three parts. In Part I, I take on the fallacious argument that hate crime laws inappropriately punish values rather than harms. In Part II, I address the mirror image fallacy that hate crime laws appropriately punish harms rather than values. Finally, in Part III, I link these two fallacies to a cluster of liberal tropes that decisionmakers and commentators use to disguise the law's morally judgmental character – a feature of our legal practices that tends to reinforce the influence of traditional social norms on law and to block progressive legal reform.

I. THE "BAD-VALUE ADDED" FALLACY

According to their opponents, hate crimes laws violate liberal principles by grounding punishment in the state's condemnation of offenders' beliefs rather than in the harms these offenders impose. Of course, all assaults impose harms that the state may legitimately seek to avert. But because an individual who is beaten or killed on account of his group affiliation is injured no more than an individual who is beaten or killed on any other account, penalty enhancements for hate crimes amount to an implicit "bad-valued added tax" on violence. The opponents typically advance this thesis to show that hate crime laws punish offenders for their "beliefs" in violation of free speech principles.[2]

As Carol Steiker has demonstrated, this argument faces a sizeable embarrassment – namely, the pervasive significance of motive in criminal law.[3] Homicide law evaluates the motive of enraged killers when it determines whether they were "adequately provoked" by their victims. The man who kills his wife because he resents her infidelity will likely be convicted only of manslaughter,[4] whereas the one who kills because he resents his wife's interference with his professional opportunities will be convicted of murder.[5] The doctrine of duress evaluates motive when it determines whether the defendant submitted to a threat that a "reasonable" person would have resisted. The woman who assists in the assault of a stranger in order to avoid being beaten herself will likely have a defense, whereas the woman who acquiesces in the assault of her own child

[2] See, e.g., James B. Jacobs and Kimberley Potter, *Hate Crimes: Criminal Law and Identity Politics* (New York: Oxford University Press, 1998), pp. 126–127; Anthony M. Dillof, "Punishing Bias: An Examination of the Theoretical Foundations of Bias Crime Statutes," *Northwestern Law Review* 91 (1997), pp. 1017–1018; Susan Gellman, "Sticks and Stones Can Put You in Jail, But Can Words Increase Your Sentence? Constitutional and Policy Dilemmas of Ethnic Intimidation Laws," *UCLA Law Review* 39 (1991), pp. 362–363.

[3] See Carol S. Steiker, "Punishing Hateful Motives: Old Wine in a New Bottle Revives Calls for Prohibition," *Michigan Law Review* 97 (1999), pp. 1863–1870.

[4] See generally Wayne R. Lafave and Austin W. Scott, Jr., *Criminal Law* (St. Paul, Minn: West Publishing Co., 2nd edn., 1986), p. 656.

[5] Cf. Commonwealth v. Carrol, 194 A.2d 911 (Pa. 1963).

in order to avoid being beaten won't.[6] The law of self-defense makes similar evaluations, treating the battered woman's desire to escape a life of degradation as an "unreasonable" ground for killing her sleeping husband,[7] but the "true man's" commitment to his honor as a "reasonable" one for killing an attacker he could easily have evaded.[8]

The law even evaluates motive in determining whether an offender deserves capital punishment. Thus, a man who kills because he wants to mutilate or to have sex with his victim's corpse can be deemed to have committed a murder that is "outrageously or wantonly vile, horrible, or inhuman."[9] For that reason, he can be sentenced to death for a homicide that would have resulted in a sentence of only life imprisonment for someone with less disgusting motives.[10]

These distinctions don't turn on the injury that the offenders impose on their victims – in each pair of cases the injuries are, roughly speaking, the same. Rather they turn on the law's assessment of the offenders' reasons for bringing those harmful consequences about. When offenders' motives show that they value the right things in the right amount (their own physical well-being over the well-being of strangers, their honor more than the lives of wrongful aggressors), the law exonerates them. When their motives show that they value the right things slightly too much relative to other things they ought to care about (e.g., fidelity more than the lives of unfaithful wives), the law mitigates their punishments. When their motives show that they value the right things far too much relative to other important things (their career advancement more than their spouses' lives, their own physical well-being more than their children's, their dignity over the lives of chronically abusive husbands), the law affords them no dispensation. And when

---

[6] *Compare* People v. Romero, 13 Cal. Rptr. 2d 332, 340 (Cal. Ct. App. 1992), *and* Morrison v. State, 546 So. 2d 102, 103 (Fla. Dist. Ct. App. 1989) *with* United States v. Webb, 747 F.2d 278, 283 (5th Cir. 1984), *and* State v. Lucero, 647 P.2d 406 (N.M. 1982).

[7] See, e.g., State v. Norman, 378 S.E.2d 8 (N.C. 1989).

[8] See generally Tennessee v. Renner, 912 S.W.2d 701 (Tenn. 1995).

[9] See, e.g., Hance v. State, 268 S.E.2d 339, 346 (Ga. 1980).

[10] See generally Dan M. Kahan, ''*The Anatomy of Disgust* in Criminal Law," *Michigan Law Review* 96 (1998), pp. 1644–1645.

their motives show they value the *wrong* things altogether (e.g., the defilement of corpses), the law punishes them all the more severely. Inconveniently for the Millian critics of hate crime penalties, the criminal law comprises a comprehensive series of bad-value added taxes.[11]

What I want to add to Steiker's critique is an account of *why* the law pays this sort of attention to offenders' motivating values. The explanation isn't that the law rejects (necessarily) the Millian position that coercion is warranted only to prevent harmful consequences rather than to sanction aberrant values. It's that the law uses offenders' motivating values to individuate harms and to measure their extent.

Start with a prosaic example. It's said that even a dog knows the difference between being tripped over and being kicked. The harms associated with being inadvertently jostled and being deliberately struck are obviously different, in nature and in magnitude, but why? The answer isn't that they invariably inflict different degrees of injury; striking is morally worse than jostling even when the latter happens to result in greater physical damage.

Obviously, we rank striking as worse than jostling because the person who strikes *intends* to harm. But why exactly does that matter? Assuming we aren't dealing with an automaton, the person who jostles another intends something too, if only to get from one place to another.

The decisive point has to do with what the striker's and jostler's respective intentions reveal about their ends. The person who deliberately strikes another values another's suffering, whereas the person who accidentally jostles another values getting wherever it is she's going. The former valuation is morally worse – it makes a bigger mistake about the objective worth of important things in the world – than the latter.

In recognizing that, we won't necessarily be committing ourselves to the conclusion that the jostler's valuations are perfectly in order. People who adequately value the worth of other individuals take precautions to avoid harming them, even inadvertently. If the circumstances surrounding the jostler's inattention suggest to us that

[11] See generally Dan M. Kahan and Martha C. Nussbaum, "Two Conceptions of Emotion in Criminal Law," *Columbia Law Review* 96 (1996), pp. 269–374.

she doesn't value others enough to take their interests into account when she forms her own plans, we'll subject her to moral condemnation. Indeed, we might subject her to legal condemnation too, in the form of civil judgment for negligence if not a criminal conviction for assault. In sum, the values that striking and jostling express are what tell us that assault and negligence are distinct harms, and that the former is worthy of greater punishment than the latter.[12]

We can deepen this point by connecting it to the expressive theory of morality associated with the work of Jean Hampton, Elizabeth Anderson, and Richard Pildes among others.[13] According to this account, social norms define how persons (and communities) who value particular goods – whether the welfare of other persons, their own honor or dignity, or the beauty of the natural environment – should behave. Against the background of these norms, actions express attitudes toward these goods.

When we morally evaluate a person's actions, we consider not only the desirability of the consequences they produce but also the appropriateness of the attitudes they express. Imagine an "all terrain vehicle" enthusiast who tried to justify the damage her hobby did to sand dunes by contributing $10,000 a year to the Sierra Club. Even if we were convinced that the contribution enabled that organization to repair more damage than the person had caused, we'd still judge the ATV enthusiast lacking in respect for the environment. Appropriately valuing the environment is inconsistent with taking pleasure in a form of recreation notorious for degrading it.

---

[12] See Richard H. Pildes and Richard G. Niemi, "Expressive Harms, 'Bizarre Districts,' and Voting Rights: Evaluating Election-District Appearances after Shaw v. Reno," *Michigan Law Review* 92 (1993), p. 510.

[13] See Jean Hampton, "The Retributive Idea," in Jeffrie G. Murphy and Jean Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988) pp. 111–161 [hereinafter Hampton, Retributive Idea]; Jean Hampton, "An Expressive Theory of Retribution," in W. Cragg (ed.), *Retributivism and Its Critics* (Stuttgart: Franz Stelner Verlag, 1992) [hereinafter, Hampton, *Expressive Theory of Retribution*]; Elizabeth Anderson, *Value in Ethics and Economics* (1993); Elizabeth S. Anderson and Richard H. Pildes, "Expressive Theories of Law: A General Restatement," *University of Pennsylvania Law Review* 148 (2000), pp. 1503–1575; Richard H. Pildes, "The Unintended Cultural Consequences of Public Policy", *Michigan Law Review* 89 (1991), pp. 936–978; Dan M. Kahan, "What Do Alternative Sanctions Mean?," *University of Chicago Law Review* 63 (1996), pp. 591–653.

By the same token, what moves us to condemn an actor for harming another isn't the simple perception that her actions have diminished another person's welfare, but rather the judgment that her actions express too low a valuation of the other person's worth relative to the actor's own ends. "[P]eople who believe their purposes warrant them in taking another's wallet, or another's savings, or another's life," Hampton writes, "are people who believe their victims are not worth enough to require better treatment."[14] "Indeed, they are convinced enough about the importance of their own purposes – and thus of their own importance – to regard their behavior as permissible with respect to these others."[15] Punishing the offender – in the appropriate way and to the appropriate degree – is necessary to show that we, in contrast, do value the victim's worth appropriately relative to the wrongdoer's interests and goals.[16]

This account helps to explain why the law sees only certain types of harmful consequences as worthy of redress. A competitor's marketing of a superior product, for example, can harm a merchant financially as much as the theft of her goods. The reason that theft but not competition merits punishment is that theft expresses – in a way that competition (ordinarily) does not – a false assessment of the merchant's worth. Against the background of social norms, the thief's behavior conveys to the merchant and to others that she views the merchant's interests as unworthy of her respect.[17]

The expressive theory also explains the law's individuation and grading of harms. The distinction between negligence and assault is one example. Another is the distinction between assault and rape. The reason we distinguish rape from assault and condemn it more severely isn't that rape invariably inflicts greater physical injury. What makes rape distinctive, and distinctively worse, is the greater contempt it evinces for its victim's agency. In our culture, we recognize the power to control one's own sexuality as a central entailment of the respect we owe individuals as autonomous, self-determining beings. For that reason, few actions convey one's subordination to

[14] Hampton, *Expressive Theory of Retribution*, *surpa*, note 13, p. 8.

[15] *Id*.

[16] See *id*., pp. 12–13; Hampton, *Retributive Idea*, *surpa*, note 13, p. 130.

[17] See Jeffrie G. Murphy, "Forgiveness and Resentment," in Murphy and Hampton, *surpa*, note 13, p. 25; Jean Hampton, "Forgiveness, Resenment and Hatred," in Murphy and Hampton, *surpa*, note 13, pp. 43–44.

another as forcefully as being made an unwilling object of that person's sexual gratification. However bad it is to assault someone, the offender who rapes makes an even bigger mistake about the worth of his victim relative to his own ends and desires.[18]

The expressive view also makes it easier to understand the critical reliance that the law makes upon doctrines that evaluate the quality of offenders' emotions. From premeditation to adequate provocation, from self-defense to duress, from insanity to the voluntary act requirement, substantive criminal law views emotions not as unthinking impulses, but rather as judgments of value. What does and doesn't enrage a person, terrify her, and even disgust her, all tell us what she cares about. If the law thinks the valuations expressed in an offender's emotions are appropriate, then the offender's passion is treated as grounds for mitigation or exculpation. If it thinks those valuations are inappropriate, then her passions become a basis for withholding mitigation or even for increasing punishment.[19]

Hate crime laws assess the values expressed by an offenders' actions in exactly the same way as the rest of criminal law. In our society, individuals tend to construct their identities around their ethnic and religious affiliations, their genders, and their sexual orientations. An individual who assaults or kills another on account of one of these characteristics, then, show us that he enjoys not only the suffering of another human being, but also the experience of domination and mastery associated with denigrating something that the victim and others regard as essential to their selves. By imposing greater punishment on those offenders, hate crime laws say that society regards the harms they impose as different from and worse than the harms inflicted by those who assault or kill for other reasons. Like assaulters relative to negligent jostlers, and rapists relative to assaulters, hate criminals relative to other types of violent criminals are – in the law's eyes – making an even bigger

[18] *Cf*. People v. Liberta, 474 N.E.2d 567, 574–575 (N.Y. 1984) (holding that "marital rape exception" violates Equal Protection Clause and rejecting argument that ordinary assault prosecution is sufficient when husband forces sex on wife: "The fact that rape statutes exist . . . is a recognition that the harm caused by a forcible rape is different, and more severe, than the harm caused by an ordinary assault. Short of homicide, [rape] is the 'ultimate violation of self'" (quoting Coker v. Georgia, 433 U.S. 584, 597 (1977)).

[19] See Kahan and Nussbaum, *surpa*, note 11.

mistake about the objective worth of their victims. Of course, the expressive parallel between hate crime laws and the rest of criminal law doesn't by itself prove that such laws are a good idea. The moral judgments that the law makes when it uses offenders' values to individuate and measure harms can themselves be morally judged. Some have argued, for example, that the law is wrong to view the true man's valuation of honor or the cuckold's valuation of fidelity as any better than the valuations expressed by persons who kill for other reasons.[20] By the same token, opponents of hate crime laws can argue that the law is wrong to view the values expressed by hate crimes as any worse than those expressed by other types of violence. What the opponents of hate crime laws can't persuasively argue, however, is that hate crime laws are a bad idea because they take values into account in a way that the rest of the law does not.

II. THE "GREATER HARM" FALLACY

Most proponents of hate crime laws accept the Millian premise that such laws must be justified on the basis of the harms as opposed to the state's aversion to hate criminals' values. Penalty enhancements are warranted, they maintain, because assaults motivated by group animus do in fact impose greater individual and societal harms than other types of assaults.[21] Responding to the First Amendment challenge considered in *Wisconsin* v. *Mitchell*,[22] defenders of hate crime laws argued that "bias-motivated crimes are more likely to provoke retaliatory crimes, inflict distinct emotional harms on their victims, and incite community unrest."[23] The Supreme Court accepted this contention as "an adequate explanation for [the state's] penalty-

[20] See Jeremy Horder, *Provocation and Responsibility* (New York: Oxford University Press, 1992), p. 39; Donna K. Coker, "Heat of Passion and Wife Killing: Men Who Batter/Men Who Kill," *Southern Californa Review of Law and Women's Studies* 2 (1992), p. 79; Kahan and Nussbaum, *surpa*, note 11, p. 365.

[21] See, e.g., Frederick M. Lawrence, *Punishing Hate: Bias Crimes under American Law* (Cambridge: Harvard University Press, 1999), p. 103.

[22] 508 U.S. 476 (1993).

[23] *Id*., p. 488; see also Lawrence, *surpa*, note 21, pp. 40–42.

enhancement provision over and above mere disagreement with offenders' beliefs or biases."[24]

The critics of hate crime laws challenge the factual basis of this "greater harm" argument. There's no evidence, they contend, to support the assertion that hate crime victims suffer greater emotional distress (an argument, in any case, that wouldn't apply to penalty enhancements for homicides), or that third parties are more likely to respond violently to hate crimes than other types of offenses.[25] Moreover, because the sponsors of hate crime legislation typically say that such laws would send a worthwhile "message" even if they didn't result in greater deterrence of bigoted crimes, critics also question the sincerity of the greater harm argument.[26]

The real problem with the greater harm argument, however, is not empirical but conceptual. Even assuming that hate crime penalty enhancements are geared to deterring greater harms, those laws can't be justified independently of the state's aversion to hate criminals' values. For as is true of the harms redressed by the criminal law generally, the "greater harms" that hate crime laws are said to deter are constructed by aversions to the values that such crimes express.

Consider the greater harms that hate crimes are said to inflict on victims. It's perfectly plausible to think that a person who is singled out for a violent assault because his attacker despises his race, religion, or sexual orientation will experience greater psychic trauma than someone attacked, say, because his attacker wants to steal his money. In our society, many individuals view their group affiliations as more central to their identities than their wealth; they

[24] 508 U.S. at 488. Defenders of hate crime laws also maintain that offenders who engage in violence because of group animus are more culpable than other violent offenders. The basis for this argument, however, is the conclusion that individuals who knowingly or purposefully inflict greater harms are more culpable than those who knowingly or purposefully inflict lesser ones. See, e.g., Lawrence, *surpa*, note 21, pp. 60–61. Accordingly, the "greater culpability" argument presupposes the "greater harm" argument.

[25] See Jacobs and Potter, *surpa*, note 2, pp. 81–88; cf. Alon Harel and Gideon Parchomovsky, "On Hate and Equality," *Yale Law Journal*, 109 (1999), pp. 514–515 (arguing that "greater harm to victims" is empirically unsupported in course of defending hate crime laws on other grounds).

[26] See, e.g., *id*. at 65.

are thus likely to experience attacks motivated by their affiliations as greater transgressions of their dignity. But if that's what's going on, the "distinct emotional harms" associated with hate crimes aren't genuinely independent of aversion to hate criminals' values. On the contrary, their aversion to their attackers' animus toward their group identities is precisely what causes victims to experience these emotional harms.[27]

Aversion to hate criminals' values plays just as clear a role in the greater "societal harms" that hate crimes are alleged to inflict. If hate crimes are indeed "more likely to provoke retaliatory crimes . . . and incite community unrest,"[28] that's because third parties, too, judge hate offenders' motives to be more reprehensible – and thus more worthy of a violent response – than those of other offenders. Sometimes hate crime law defenders say (plausibly, in my view) that bias motivated assaults are perceived as an attack by all individuals who share the group affiliations of the victim.[29] But if that's so, what's distressing those third parties is *their* reaction to the low value that hate crime offenders assign to members of their group.

Once she recognizes this relationship between harms and values, an orthodox Millian liberal shouldn't see any difference between a statute that is predicated on deterring those harms and one that is expressly predicated on repudiation of hate criminals' values. It wouldn't be a defense to a discrimination claim for an employer to say she refused to hire African Americans not because she was prejudiced, but only because her *customers* doubted the competence of African–American employees. In that case, the law would treat the responsiveness of the employer to her customers as simply the mechanism by which racial prejudice was constraining African–Americans' job prospects.[30] By the same token, it's no defense to the Millian objection for state legislators to explain that they are enacting hate crime laws not because *they* happen to disagree with hate offenders' beliefs but only because their *constituents* experience distinctive harms from those crimes. Because those harms are

[27] *Cf*. Lawrence, *surpa*, note 21, p. 62.

[28] 508 U.S. at 488.

[29] See, e.g., Lawrence, *surpa*, note 21, pp. 42, 63.

[30] *Cf*. Diaz v Pan American World Airways, Inc., 442 F.2d 385 (5th Cir. 1971) (holding that passengers' preference for female flight attendants did not immunize airline from a sex discrimination claim by prospective male flight attendants).

constructed by aversions to hate criminals' values, a Millian liberal should treat the responsiveness of legislators to their constituents as simply the mechanism by which the law credits those aversions as a ground for punishment.

Indeed, Millian liberals have had no problem seeing aversions to values at the core of asserted societal harm arguments in other settings. The traditional justification for punishing homosexuality is the offensiveness of same-sex desire pure and simple. In his famous critique of the Wolfenden Report, Lord Devlin tried to strengthen the traditional justification, arguing that the ban on homosexuality reflected the danger that social order would dissolve were the law to disregard its citizens' most intensely held moral convictions.[31] H.L.A. Hart, Ronald Dworkin, and other Millian liberals wasted no time in pointing out that Devlin's argument added nothing, or at least nothing of consequence, to the traditional argument from offense.[32] Even assuming that Devlin's alarming conjectures were empirically sound, the disorder he anticipated would itself have originated in the disappointment of citizens' demand that offensive behavior be prohibited; in punishing homosexuality to avoid the "harm" of social disorder, then, the law would still have been treating aversion to same-sex desire as a legitimate ground for coercion. The relationship between social disorder and aversions in Devlin's argument is indistinguishable from the relationship between social disorder and aversions in the "greater harm" argument that the Supreme Court accepted in *Wisconsin* v. *Mitchell*.

This parallel certainly doesn't mean that supporters of hate crime laws must climb into bed with Devlin when it comes to the criminalization of homosexuality. But it does mean that if one wants to defend hate crime laws at the same time that one criticizes homosexual sodomy laws, one must explain why the law is *right* to credit aversions to hate criminals' values and *wrong* to credit aversions to same-sex desire.

---

[31] See Patrick Devlin, *The Enforcement of Morals* (New York: Oxford University Press, 1965).

[32] See H.L.A. Hart, *Law, Liberty, and Morality* (New York: Vintage Books, 1963); H.L.A. Hart, "Social Solidarity and the Enforcement of Morality," *University of Chicago Law Review* 35 (1967), pp. 1–13; Ronald Dworkin, "Lord Devlin and the Enforcement of Morals," *Yale Law Journal*,] 75 (1966), p. 986.

III. THE CONSERVATIVE BIAS OF LIBERAL ANTIMORALISM

The decisive question in the hate crime debate can't be whether hate crime laws punish values or harms. Without taking account of the values that offenders' actions express, the law wouldn't be able to individuate harms or measure their severity. The only way to determine whether hate crime laws are appropriate or not, then, is to assess the moral appraisal *they* express head on: is it true that persons who engage in violence because of group animus devalue their victims in a manner that is distinct from and more reprehensible than those who engage in violence for myriad other reasons?

The remainder of this essay, however, isn't designed to answer this question but rather to motivate the insistence that the protagonists in the hate crimes debate expressly attend to it. Far more than analytical clarity is at stake. By framing their disagreement in a manner that elides the need to take a stance on the values of hate crime offenders, both sides reinforce the dominance of a demoralized form of legal discourse that ultimately makes the law more congenial to traditional social norms and more resistant to progressive reform.

The language of substantive law doctrines tends to obscure the morally judgmental stance of the law toward offenders' values. To the untrained eye, homicide doctrines like "premeditation,"[33] "heat of passion,"[34] and "extreme mental or emotional disturbance"[35] seem to make the severity of punishment turn on the force rather than on moral quality of defendant's motivating passions. So too do defenses like duress, which is said to excuse when "the will of the accused has been overborne by threats,"[36] insanity, which excuses offenders who are subject to "irresistible impulses,"[37] and

---

[33] See Wayne R. LaFave and Austin W. Scott, *surpa*, note 4, §7.7.

[34] *Id*., §7.10.

[35] Model Penal Code §210.3(1)(b).

[36] E.g., Regina v. Hudson, [1971] 2 All E.R. 244, 246.

[37] See Abraham Goldstein, The Insanity Defense (New Haven: Yale University Press, 1967).

self-defense, which courts sometimes describe as a concession to humans' "primal impulse" of "self-preservation."[38]

This impression is reinforced by the two dominant scholarly theories of criminal law. *Voluntarism*, which derives from Kantian moral philosophy, treats punishment as justified if, and to the extent that, the offender's behavior stems from choice. *Consequentialism*, which derives from utilitarian theory, views punishment as warranted if, and to the extent that, visiting suffering on the offender is necessary to promote favored states of affairs. The former theory cares about an offender's motivating passions only to the extent that they interfere with volition, and the latter only to the extent that they impel the offender to act in ways that the law wants to encourage or discourage. Neither theory purports to assign any independent normative significance to the valuations that offenders' motivations express.[39]

But precisely because they ignore the expressive logic of the criminal law, the naïve reading as well as the conventional theories are in fact inadequate. Homicide gradations and criminal law defenses are subject to various qualitative delimiters – "adequate provocation," "reasonable person" requirements, and even "disease," a concept the law defines in distinctively moral rather than medical terms. These delimiters, it can be shown, withhold mitigation to offenders whose passions and impulses express what judges and juries see as inappropriate values, no matter how destructive those passions are of offenders' capacities to choose.[40] It's true in some sense that these doctrines tailor punishment to the tendency of offenders to promote or frustrate favored states of affairs. But upon investigation, it turns out that the the state of affairs that the law favors are derivative of the moral appraisals decision-

[38] E.g., State v. Norman, 378 S.E.2d 8, 12–13 (N.C. 1989); see also Thomas Hobbes, Leviathan (New York: Penguin Classics, 1985) (1651), pp. 345–346 ("If a man by the terrour of present death, be compelled to doe a fact against the Law, he is totally Excused; because no Law can oblige a man to abandon his own preservation. . . . Nature . . . compels him to the fact.").

[39] See Kahan and Nussbaum, *surpa*, note 11, pp. 301–304. Hurd's contribution to this issue neatly encapsulates these conventional – but deeply flawed – accounts of how emotions and like dispositions figure in criminal law doctrine. See Heidi Hurd, *Why Liberals Should Hate Hate Crime Legislation*, L. and Phil.

[40] See Kahan and Nussbaum, *surpa*, note 11, pp. 301–346.

makers make of the values expressed in offenders' motivating passions.[41]

It's possible, in short, to dispel the doctrinal obscurity and misdirected theorizing that conceal the law's moralistic core. But it takes a large amount of philosophical labor to do so. Why does our legal order make so sizeable an investment in constructing a morally sanitized idiom to describe and justify its basic operations?

The answer is the influnce of a particular brand of liberalism in our legal and political culture. Accepting the fact of permanent moral dissensus, liberalism so understood enjoins us to justify our positions on public issues in terms acceptable to those of diverse cultural and moral persuasions.[42] This directive is reinforced by social norms that expose ordinary citizens to censure for public moralizing, and by strategic incentives that discourage (most) public officials from resorting to culturally and morally divisive rhetoric. As a result of these dynamics, participants in the legal culture converge on forms of discourse that abstract from and elide the morally partisan judgments that animate their positions on contentious issues. The mechanistic idiom that pervades substantive criminal law doctrine and the de-moralized abstractions that inform voluntarism and consequentialism satisfy this function.[43]

So do the terms in which the hate crimes debate is conventionally framed. Mill's "harm" principle is, of course, a mainstay of liberal moral theory. It tests whether coercion does in fact have a justification independent of the ambition to impose a partisan conception of the good on those who disagree with it. By invoking Mill's harm principle, the critics of hate crime laws can explain their opposition without directly contesting the moral and ideological commitments of these laws' supporters. By purporting to conform these provisions to the harm principle, defenders of hate crime laws needn't acknowledge their ambition to use the expressive capital of the law to underwrite their struggle against hierarchical social norms.

---

[41] See *id*. at 312, 335–336.

[42] See generally John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993); Bruce A. Ackerman, *Social Justice and the Liberal State* (New Haven: Yale University Press, 1980); Amy Gutmann and Dennis Thompson, *Democracy and Disagreement* (Cambridge: Belknap Press, 1996).

[43] See generally Dan M. Kahan, "The Secret Ambition of Deterrence," *Harvard Law Review* 113 (1999), pp. 413–500.

Whatever might be said in favor of the power of liberal discourse strategies to mute political conflict in law in general,[44] its power to obscure the morally judgmental character of criminal law in particular should be regarded as pernicious. For in that context, the irrepressible tendency of such strategies is to stack the rhetorical deck of the law in favor of traditionally hierarchical social norms and against progressive egalitarian ones.

The mechanistic veneer of criminal law doctrines, voluntarist and consequentialist rationalizations of them, and arguments founded in the Millian harm principle all falsely purport to justify coercion without resort to contestable moral valuations. To determine whether a person "reasonably" submitted to her anger or fear, we must first contentiously assess whether it was reasonable, morally, for her to value the good that was endangered by the provoking or threatening circumstance. To determine whether punishing an individual will promote desired states of affairs, we must first make contentious judgments about what state of affairs is desirable. To identify and evaluate the harms associated with acts of violence, we must first identify the values that motivate individuals to engage in them.

The only situation in which liberal tropes will succeed in concealing decisionmakers' contestable moral judgments is when those judgments are rooted in social norms that have traditionally enjoyed widespread consensus. Our perceptions of "choice," "dangerousness," "harm" and the like are all socially constructed. But so long as the law doesn't diverge too sharply from traditional social norms, the work that those norms do in constructing the apprehension of these phenomena will be essentially invisible. In a society that has traditionally valued patriarchal conceptions of honor, it will seem perfectly natural for "a man of good moral character[,] . . . highly respected in his community, [and] having regard for his duties as a husband and the virtue of women," to be "shocked" to the point of "temporary insanity" upon "learning

[44] For defenses, see David A. Strauss, "Legal Arguments and the Overlapping Consensus" (unpublished manuscript, July 12, 1998); Cass R. Sunstein, *Legal Reasoning and Political Conflict* (New York: Oxford University Press, 1996), pp. 35–47.

of the immorality [i.e., infidelity] of his wife."[45] In one that has traditionally been repulsed by homosexuality, it will seem perfectly obvious and uncontroversial to describe the killing of a gay person by a man he solicited at a bar as a "one-time tragedy" committed by an otherwise peaceful individual who the decisionmaker can be "confident . . . w[ill] not kill again."[46]

The situation is strikingly different, however, when legal reformers advocate doctrinal reforms founded on emerging but still highly contested norms. At that point, the evaluative work that those norms and associated moral evaluations are doing becomes transparent.[47] And against the background of the de-moralized liberal tropes that decisionmakers and commentators use to justify the law, the law's receptivity to these nontraditional moral claims will always seem transparently illegitimate.

Victoria Nourse makes this point to explain the influence of the so-called "abuse excuse" critique.[48] Commentators advance this critique against the law's growing receptivity to forcible self-help by the victims of chronic domestic violence.[49] The scientific evidence underlying "battered woman syndrome" and like conditions, these commentators argue, is far too weak to support the conclusion that individuals who resort to violence in such circumstances are suffering from genuine volitional impairment; courts accept such evidence nonetheless, the critics claim, only because they sympathize with the ideological convictions that inspire these doctrinal reforms. But as Nourse notes, the law has *always* indulged facile claims of volitional impairment to excuse the virtuous outlaw.

---

[45] Hamilton v. State, 244 P.2d 328, 335 (Okla. Crim. App. 1952).

[46] See Steven Hunt, "Victim's Family, Gays Say Killer Got Off Too Easy," *Salt Lake Tribune, Aug*. 16, 1997, p. C7; "Judge Draws Protest After Cutting Sentence of Gay Man's Killer," *N.Y. Times*, Aug. 17, 1994, p. A15.

[47] See Steiker, *surpa*, note 3, p. 1870.

[48] See Victoria Nourse, "The New Normativity: The Abuse Excuse and the Resurgence of Judgment in the Criminal Law," *Stanford Law Review* 50 (1998), pp. 1435–1470.

[49] See, e.g., Alan M. Dershowitz, *The Abuse Excuse: and Other Cop-outs, Sob stories, and Evasions of Responsibility* (Boston: Little, Brown, 1994). See also Kahan and Nussbaum, *surpa* note 11, at 349; James Q. Wilson, *Moral Judgment: Does the Abuse Excuse Threaten Our Legal System?* (New York: BasicBooks, 1997).

The only thing that makes the "abuse excuse" epithet resonate is that the identity of the virtuous outlaw has changed as traditional hierarchical norms have come under attack from new, egalitarian ones: he used to be the vengeful cuckold; she's now the battered woman.[50]

The same myopia, induced by the same bias against emerging norms, is the only thing that makes the Millian critique of hate crime laws appear plausible. The critics object to the use of such laws to condemn offenders' aberrant values rather than the harms they inflict. But conceptually and practically speaking, the role that values play in defining the harm addressed by hate crime plays is no different from the role that they play in defining the diverse harms addressed by homicide gradations, rape law, capital sentencing provisions, and a host of other doctrines. Condemnation of values is more visible in hate crime laws than in these other areas only because the norms that inform hate crimes, unlike the norms that inform these other provisions, are nontraditional and contested.

So long as the official language of criminal law remains the demoralized idiom of liberalism, progressive reformers will continue to be at a disadvantage in doctrinal debates. The weakness of the "greater harm" argument attests to the futility of trying to adapt the conventional liberal tropes to progressive ends. In the hate crime debate and elsewhere, progressives should ruthlessly expose the partisan moral judgments that underlie the doctrines they want to abolish, and defend without embarrassment the ones that underlie the doctrines they propose instead.[51]

CONCLUSION

The proponents and opponents of hate crime laws share an assumption: that the legitimacy of such laws depends on whether they deter harms independent of the state's aversion to hate criminals' noxious ideologies. This shared assumption reflects the influence of orthodox liberal theory, which holds that coercion is warranted

[50] See also Kahan and Nussbaum, *surpa*, note 11, pp. 349–350.

[51] See generally Dan M. Kahan, "The Progressive Appropriation of Disgust," in Susan A. Bandes (ed.), *The Passions of Law* (New York: NYU Press, 2000).

only to prevent harm to others and not to condemn individuals for subscribing to aberrant values.

I have tried to show that the debate, so framed, is miscast. Conceptually and practically speaking, it is impossible to draw a distinction between the harms that violent criminals inflict and the values that motivate them to act. Hate crime laws do punish offenders for their aberrant values in this sense. But so do the rest of the provisions that make up criminal law. It's impossible to imagine things being otherwise.

The real question that those for and against hate crime laws should be debating is whether those laws are right to see hate criminals' values as more worthy of condemnation than those of other violent offenders. No light will be shed on this issue – or on a variety of other urgent topics – until both sides step out from behind the shadow of John Stuart Mill.

Yale Law School
P.O. Box 208215
New Haven CT 06520-8215
USA
(E-mail: dan.kahan@yale.edu)