Hate and Punishment

Antti Kauppinen

Trinity College Dublin

Final draft, June 4, 2014

For *The Journal of Interpersonal Violence*

Special issue on hate crime, edited by Christian Munthe and David Brax.

According to the view I will call Legal Expressivism (LE), neither crime nor punishment consists merely in intentionally imposing some kind of harm on another. Both also have an *expressive* aspect. In terms I will be using, they are what they are in part because they *enact* attitudes towards others – in the case of crime, some kind of disrespect, at least, and in the case of punishment, society's condemnation or reprobation. Indeed, punishment is *justified*, at least in part, because (and when) it uniquely expresses fitting condemnation or other retributive attitude. Hate or bias crimes dramatize the expressive aspect of crime, since they typically, and sometimes by design, send a message of inferiority to the victim's group and society at large. Treating the enactment of contempt and denigration towards a historically underprivileged group as an aggravating factor in sentencing may be an appropriate way to counter this message, since it reaffirms and indeed realizes the fundamental equality of all members of a democratic community.

Elements of Legal Expressivism can be found in the work of many legal thinkers.[1] Yet virtually every aspect of the story has been called into question by critics. In this paper, I will refine the LE account and respond to some of the criticisms. I will begin with the notion of an attitude and its relation to emotion, and argue that the two should not be assimilated.

[1] See in particular Joel Feinberg (1965), Jeffrie Murphy & Jean Hampton (1988), Jean Hampton (1992), Martha Nussbaum & Dan Kahan (1996), Dan Kahan (1996, 2001), Elizabeth Anderson & Richard Pildes (2000), Antony Duff (2001), and Christopher Bennett (2006).

I'll also distinguish between two ways of expressing attitudes: enactment and symbolic expression. Hate crimes, in these terms, are crimes that enact hate, contempt, or disgust towards someone on account of identity-defining group membership.

How can expressing such attitudes justify enhanced punishment? I begin by discussing appeal to a special kind of *harm* caused by hate crime and expressing attitudes in general, and argue that no kind of additional harm suffices to provide a rationale for enhanced punishment for hate crime. Instead, we need to turn to the LE framework. On this picture, wrongdoing is not merely about causing harm, but is rather a function of attitudes towards others (or, equivalently, the valuations) that our actions embody. Punishment manifests reactive attitudes, and is justified in terms of countering the evaluation implicit in the criminal's behaviour. I take up various criticisms of LE, and argue that they fail to damage the most plausible version, which appeals to the conceptual link between punishment and protecting the status of inviolability that all citizens have in a democratic community in justifying legal condemnation. This is a kind of mixed theory of punishment, since it has both backward-looking elements (punishment expresses reactive attitudes) and forward-looking ones (reactive attitudes are fitting because of their role in protecting the inviolable status of the victim).

Given this understanding of crime and punishment in general, we can see the proper place of hate crime legislation. Hate crime is a greater wrong than the parallel crime, because it enacts attitudes that pose a graver threat to the inviolability of all members of the target group. In response, the right kind of hate crime legislation reaffirms the equality of all citizens as a fundamental democratic value. It expresses the determination on the part of members of the democratic community not to allow anyone to be treated as an object of disgust, contempt, or disdain, as a lower kind of being merely in virtue of belonging to a

group that is for some reason disvalued. In so doing, it preserves the equal status of all citizens, since such status supervenes on enacting reactive attitudes towards violations.

1. Attitudes and Their Manifestations

I will begin with the notion of an attitude or sentiment, since it is crucial to any LE account. Though the term is sometimes used loosely, an attitude is distinct from an emotion or passion or affect. Rather, an attitude is a way of relating to someone or something that disposes one to have different emotions in different situations, to want certain things, to focus attention on certain things, to deliberate in certain ways, and possibly to make use of evidence in certain ways (cf. Anderson & Pildes 2000, 1509). My attitudes show how much I value someone or something. Thus, if I love someone, I will feel joy when she thrives and sadness if she suffers, I want her to do well for her own sake and am committed to doing things that further her good, I notice things that are opportunities or threats for her, I take those opportunities and threats into account in practical deliberation as reasons for action, and may be more likely to believe that she is excellent at her job than the evidence warrants. Perhaps there is a distinct feeling of love that goes with the attitude, but love itself is more than that feeling.

Moreover, in the case of an attitude like love, I welcome or endorse the reactions associated with the sentiment, rather than being in their grip. In one sense of the word, I can *identify* with the pattern of passive responses. Only once I do so, the attitude will be genuinely mine. This doesn't mean that the attitude is under direct voluntary control, but neither is it purely passive, as it is sometimes portrayed. In fact, there is plausibly an asymmetry between our control over loving someone and not loving them. We can't will love into existence, but we may be able to will it *out* of existence. That is, if thoughts about

someone give rise to no positive feeling or spontaneously focus our attention, there's nothing we can directly do to change that. That's the problem with some arranged marriages. In contrast, if I'm already in a relationship and find myself spontaneously attracted to a co-worker, say, I may well be able to deliberately set aside the feelings and refocus my attention on something else, and catch myself before falling in love.

Hate, as I understand it, is in many ways the mirror image of love. If I hate someone, I want her to do badly, whether or not it is of instrumental benefit for me. I feel bad if she does well, get easily angry with her, and may be delighted if misfortune befalls her. I tend to notice things that are opportunities or threats, and take those as reasons for action – though obviously in the opposite way to the case of love. In full-blown hate, I don't regard these dispositions as something I want to get rid of, but embrace them, perhaps even take pride in them. Again, I may not be able to will hate into existence, but still able to will it out of existence. Perhaps I realize that I'm responding to Jewish people in the same negative way, regardless of their individual qualities. Insofar as I refuse the allow these feelings and thoughts to influence my deliberation and action, I am not a full-blown anti-semite – though of course, I'm not quite innocent either, until and unless the passive responses cease.

I don't want to exaggerate the similarities between hate and love, however. One important difference between them is that hate may be entirely *impersonal*, as Thomas Brudholm (2010, 295) emphasizes – the individual characteristics of the hated person need not matter, if she is hated on account of belonging to a group. Relatedly, unlike resentment or indignation, hate does not essentially seek *punishment* of its target for what she's done. There doesn't have to be a wrong that should or could be repaid or forgiven, although hate may of course be triggered by wrongdoing as well. As Brudholm observes, even when it is a

kind of personal reactive attitude, hate targets the character of the person rather than particular actions (ibid., 306–307).

One important consequence of the general picture of attitudes I have introduced is that unlike occurrent emotions, attitudes are not *transparent*. It may be that you can't be mistaken about whether you feel joy or sadness or rage right now, but you can be mistaken about whether you love someone – or whether you hate someone. When having an attitude goes against social norms, false belief about it may even be the normal condition. In many circles, anti-semitism, for example, is very disreputable, so people who are prejudiced towards Jews may sincerely but falsely believe that they are not. Other people can be better judges of our attitudes than we are ourselves.

We can manifest or express our attitudes in two different ways. We can do so by way of using language or some other convention-based system of signs that is designed to convey our stance to others. I can express my love to my wife by telling her that I love her or by making her something she likes. As I use the term, the expression relation is non-causal, non-factive, and not necessarily intentional. Speech and symbols can express psychological states we don't actually have, as happens when we're insincere. They can also express attitudes we're not aware of having or don't mean to express. Otherwise psychoanalysts would go out of business.

The other way of expressing attitudes is acting in a way that is best made sense of by attributing the attitude to the agent. I'll talk about *enacting* the attitude in this case to contrast it with *symbolic expression*. For example, I enact my love for someone by going out of my way to comfort her when she is in need, whether or not I say anything about my feelings. I just express the love by performing the actions that it motivates me to do. Now, as I've defined enacting, it is also potentially non-factive. It may be that what makes best sense of

my behaviour is an attitude I don't actually have. It may be that what I do and the way I do it is such that any reasonable person, well-versed in my cultural context would conclude that I love someone, were they the examine my behaviour, when in fact I am entirely indifferent. Still, on my view, I have enacted love, and people who attributed the attitude to me would be justified in doing so, though mistaken about my mental states. Or, to put it differently, they would be mistaken about my actual attitudes, but not about the attitude *embodied* by my behaviour. It is a common mistake in this context to liken or assimilate this kind of expression to linguistic meaning (see e.g. Hampton 1992, Adler 2000). In enactment, convention plays at best an indirect role, and its pragmatic effects are very different.

My view belongs to what we may call the *externalist* family, since according to it the meaning of an action isn't determined by what goes on inside the agent's (or any particular interpreter's) mind. Instead, "to grasp the expressive meaning of an act, we try to make sense of it by fitting it into an interpretive context ... [A] proposed interpretation must make sense in light of the community's other practices, its history, and shared meanings." (Anderson & Pildes 2000, 1525). Think about a case of everyday jealousy. A friend of mine is a very good singer and musician. Recently, he went to a wedding party with his wife, who urged him to grab an acoustic guitar lying around. When he reluctantly did so, some twenty-something girls gathered around him, showering him with requests and adulation. He complied with the requests and welcomed the adulation, with the result that his wife subsequently accused him of flirting with the girls. My friend felt unfairly judged, since he wasn't in fact interested in the young women (or so he says). However, his wife had a point, too: his behaviour did reasonably express an interest. Without going too deep into what makes for a reasonable interpretation, a good starting point is asking oneself about the goals that an informed and rational agent in the relevant cultural context would pursue by engaging in the observed kind

of behaviour. There will inevitably be indeterminacy here, but that is the kind of fact of life that must often be acknowledged in the context of law.

One relevant issue in this area will be whether *groups* of people, such as legislatures or courts, can have or express attitudes above and beyond the attitudes of the individuals that comprise them. The philosophical pendulum has swung in the direction of the legitimacy of attributing group attitudes (see e.g. Tuomela 2007; List & Pettit 2011). Following this tradition, I will take it as given that the attitudes of a collective (such as a state) are determined by the goals, assumptions, and inferences that enough of its relevant members jointly accept as the basis of collective action. As members of a group, people can take as given premises that they wouldn't make use of in their personal practical or theoretical reasoning, which means that we can attribute to a group goals or beliefs that a minority of members, or no one, has. For example, military planners might agree to design an armament program starting from the assumption that the Chinese will have a new fighter plane in 2016, even if few of them personally think they will be ready yet.

Which attitudes should be ascribe to a group, then? I will take Blackburn's Credibility principle as a starting point:

> A group may be said to have been committed to a belief (goal, principle) if there is no way – no credible way – that the group could rationally sustain their open affirmations were they not also prepared to stand by the belief (goal, principle). (2010, 81)

The most straightforward way a group can commit to a goal is announcing, after going through the procedures its members consider as authoritative, that it has decided to do such-and-such. Thus, the European Central Bank may make it known that it will lend money to Greece. But Credibility also leaves room for groups having beliefs and goals (and indeed attitudes) that are not directly affirmed by any of the relevant spokespeople. The ECB might be committed to preferring a rise in unemployment to a rise in inflation, if it couldn't

credibly deny such a preference, given the pattern of past and present decisions, even if it is never openly affirmed by anyone.

2. Hate Crime and the Varieties of Harm

No one, I think, would deny that some of the attitudes we have, such as hate or bias, are *morally* inappropriate. But should having or manifesting such attitudes be the business of the *law*, as advocates of hate crime and hate speech legislation want? Isn't law concerned with our external behaviour rather than motives, as Kant, among others, argued? Is it right to punish one offender more strictly than another, if the only difference between them is in the quality of their motives?

The commonly accepted starting point for answering these questions is the principle of proportionality: deserved punishment for a crime is a function of both the degree of wrongdoing and the degree of culpability or responsibility (see e.g. Al-Hakim and Dimmock 2012, 579ff). If hate crimes deserve a stricter punishment, they must either amount to a greater wrong or involve greater culpability. The latter, plausibly, is primarily determined by factors such as intentionality, presence of coercion or invincible ignorance, negligence, and so on. With respect to them, there is unlikely to be a systematic difference between hate crimes and parallel crimes. Much of the discussion about hate crime, including this paper, thus focuses on the difference in wrongdoing.

What determines the degree of wrongdoing? One simple answer is that wrongdoing is at least in part a function of the degree of *harm* to others that an action causes. Perhaps rape is worse than sexual harassment, because it is more harmful to the victim. Along this line, some have defended stricter punishment for hate crimes on the grounds that they cause *greater harm* to the victim (Lawrence 1999; Iganski 2001; Wisconsin v. Mitchell, 508 U.S.

476 (1993)). For example, there is some evidence that the psychological and physical harms of hate crime are likely to be more serious or more widespread than those of parallel crimes.

However, Heidi Hurd and Michael Moore (2004) argue convincingly that enhanced punishment for hate crime cannot in fact be defended on the grounds that such crimes cause greater physical, psychological, collateral, or social harm than otherwise motivated crimes. This, they rightly point out, is a contingent matter. If a greed-motivated crime, say, causes the same harm, why should it be punished more lightly? Motivation here is only a poor proxy for what really makes for the degree of seriousness of crime, namely harm (Hurd & Moore 2004, 1086).[2] Similarly, if it is *morally worse* to harm someone who is in a vulnerable position than someone who isn't, this doesn't justify differential treatment of a perpetrator who targets someone vulnerable because of her race, for example – it's the vulnerability rather than the sort of characteristics that hate crime legislation involves that justifies the difference (ibid., 1097). Conversely, as Anderson and Pildes (2000) point out, thick-skinned victims of hate crime may be psychologically unaffected by the hate aspect. This leads to the ridiculous consequence that "only the thin-skinned and psychologically fragile are entitled to be treated with dignity" (Anderson & Pildes 2000, 1543). So it's no good to appeal to greater conventional harm to justify enhanced punishment for hate crime.

Many defenders of hate crime legislation recognize the difficulties with appealing to ordinary kinds of harm, and consequently introduce a new concept of distinctive *expressive* harm. This, then, provides a rationale for enhanced punishment. As Dan Kahan puts it, "By imposing greater punishment on [hate] offenders, hate crime laws say that society regards

[2] A reviewer for this journal pointed out that one could nevertheless defend hate crime legislation on grounds of posing a *risk* of greater harm, parallel to drunk driving, for example. I think there is something to this suggestion. But I believe that the reason why merely imposing a risk on others (even when no harm actually results) can be legally punishable wrongdoing is ultimately that doing so expresses an inappropriate attitude towards them. Thus the wrongdoing involved in imposing risks is not to be understood in terms of harm but rather the Expressive Wrongdoing view sketched in Section 3 below.

the harms they impose as different from and worse than the harms inflicted by those who assault or kill for other reasons." (Kahan 2001, 182) Kahan does not do much to explicate the different kind of harm – indeed, his considered view is better construed as a version of the Wrong Valuation account I discuss below. For a full discussion of what makes for expressive harm, we need to turn to Elizabeth Anderson and Richard Pildes (2000).

Anderson and Pildes' argument proceeds as follows. They begin with the claim that "A person suffers expressive harm when she is treated according to principles that express negative or inappropriate attitudes toward her." (Anderson & Pildes 2000, 1527) This definition does not by itself say what makes expression of inappropriate attitudes a non-psychological harm for a person. Their answer draws on a link between expression of attitudes and social relationships: "The communication of attitudes creates social relationships by establishing shared understandings of the attitudes that will govern the interactions of the parties." (1528) Given this, "Communications can expressively harm people by creating or changing the social relationships in which the addressees stand to the communicator" (1529), where these social relationships include "friendship and enmity, collegiality and rivalry, and superior and inferior caste status" (1530). So the argument seems to have three premises:

1) Social relationships supervene on a shared understanding of the attitudes that govern mutual interactions.
2) Shared understandings of interaction-governing attitudes result from actions that express attitudes.
3) Standing in certain social relationships is bad for a person, regardless of how they feel about it.
4) Hence, actions that express attitudes can harm a person via changing the relationships in which they stand to others.

The first premise concerns the nature of social relationships. It says that what makes us colleagues, for example, is that we mutually understand that we are to relate to each other in

certain ways, and, perhaps, not in other ways – for example, we share an understanding that we won't hate or love each other. (If we do, we don't do so *qua* colleagues.) The second premise says that the shared understandings are created by expressing attitudes. To be sure, the understanding that governs our relations may not be the result of anything you and I specifically did. In most cases, we inherit the conception of what goes with what role from our cultural and historical background. But our expressive actions reaffirm and possibly shape the understanding we have, and perhaps the way we acquire the tradition is by way of expressive action. At least, this is not wholly implausible. The last bit, then, is that some social relationships are as such bad for us to stand in. Which relationships? Anderson and Pildes don't say, but their example, not of hate crime but of expressive harm of state action, illustrates what they mean:

> Racial segregation sends the message that blacks are untouchable, a kind of social pollutant from which "pure" whites must be protected. For the communicative goal to be realized, its meaning must be acknowledged. This does not mean that the addressees must believe, approve of, or accept the message. They simply have to understand it. Once people share an understanding that segregation laws express contempt for blacks, these laws constitute blacks as an "untouchable," stigmatized caste. (Anderson & Pildes 2000, 1528)

So on this view, the expressive harm of hate crime, independently of its psychological effect on people, consists in its manifesting attitudes toward the victim that place the offender and the victim in an undesirable and harmful relationship in which one dominates the other. If that is the case, enhanced punishment is justified on the basis of greater wrongdoing than in the parallel crime.

As Section 4 will show, I find much that is congenial about this picture. However, I don't believe it will quite suffice. Some social relationships may be all about how we feel about each other: what makes us lovers is that we love each other. But other relationships involve more than a shared understanding of interaction-governing attitudes, and some of those are highly relevant in the context of hate crime. Relationships of domination and

subordination and oppression, in particular, may exist in the absence of such understanding. I may think that I am the social or legal equal of someone above me in a hierarchy – and in some cases, they may think the same way – but sadly, that isn't sufficient to make us equals. Relationships of relative power do not reduce to how we understand.

The problem for Anderson and Pildes is that to claim that simply *understanding* that the behaviour of others expresses contempt towards me as a disabled person, for example, while nevertheless not accepting that it is in any way merited, suffices to make me stigmatized or disadvantaged makes little sense. Instead, being dominated or subordinated is a *status* that one may have. As I will argue in the next section, status does not supervise just on attitudes that others express towards me, but on how society, and the law in particular, responds to those attitudes. What matters for status is not just symbolic communication but concrete enactment through punishment.

3. Punishment, Expression, and Inviolability

If the arguments of the previous section are correct, it is not greater or different kind of *harm* that makes hate crime deserve enhanced punishment. A different understanding of both crime and punishment is called for. This is what Legal Expressivism promises. Proponents of this type of view emphasize that both crime and punishment have an attitudinal or communicative dimension. It is common for proponents to begin with the observation that merely causing harm or loss to another, even if done intentionally, does not suffice for punishable crime. Instead, we need to look at the attitudes expressed by the perpetrator. Many emphasize that these attitudes embody a *mistaken evaluation* of the victim's worth. As Dan Kahan puts it,

> 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. (Kahan 2001, 181; cf. Kahan and Nussbaum 1996, 351–2)

Kahan draws on the work of Jean Hampton, who argues that treating someone in a way that is precluded by her value or represents her as less valuable than it is "diminishes" her value, and simultaneously represents the agent as elevated with respect to the victim (Hampton 1992, 1672–7; cf. Murphy & Hampton 1988, 44-5, 124). Note that this is not the same as the victim *feeling* demeaned or insulted, or losing self-respect. Nor does it require that the perpetrator *intends* to diminish the value of the victim – that's not what a pickpocket has in mind. The wrongdoing involved in crime is a function of the valuation it expresses rather than harm alone. Call this the *Wrong Valuation* account of wrongdoing.

Let me be explicit that on this view, wrongdoing is a matter of the attitude enacted by the action, or what some call its social meaning. Since I've endorsed an externalist construal of social meaning, this view is a target of the following objection raised by Hurd and Moore:

> [Externalist] construals of social meaning … 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). (Hurd & Moore 2004, 1107)

Hurd and Moore argue that such appearance could not possibly justify punishment, especially when false. However, their reading misconstrues the externalist view. There *is* a fact about the deed – for example, that it is disrespectful of someone. This fact is nothing other than the fact that it would be reasonably taken to be so by someone aware of the features of the action and context. Blameworthiness is a separate issue. For Legal Expressivism, what makes the action as potential reason for blame in the first place is that it (objectively) expresses bad attitude or wrong valuation.[3] Someone who performs a hateful

[3] Hampton holds a similar view: "[T]he message of the action and the actuality of what it accomplishes is not only something that we understand apart from the victim's reaction to it, but also

act, for example, without realizing it is such (and thus not acting out of hate) may be excused on ordinary grounds – the act wasn't intentional under the description, the agent was invincibly ignorant of the social meaning of the act, and so on.

The next step is that *punishment* expresses what Peter Strawson (1962) labeled *reactive attitudes* – attitudes towards the attitudes or quality of the will that the agent has expressed by his action. Such attitudes can meaningfully be attributed to groups on the basis of goals and assumptions that their representatives could not credibly disavow, given their behaviour. In the case of punishment, the reactive attitudes are those of the community through authoritative representatives. A classic formulation is given by Joel Feinberg's definition of punishment, though Feinberg gives it a more conventionalist twist than necessary:

> Punishment is a conventional device for the expression of attitudes of resentment and indignation, and of judgments of disapproval and reprobation, either on the part of the punishing authority himself or of those "in whose name" the punishment is inflicted. (Feinberg 1965, 400)

For Legal Expressivists, expressing reactive attitudes is what *makes* something a punishment. In Antony Duff's words, punishment is "a mode of moral communication with offenders that seeks to persuade them to repent their crimes, to reform themselves, and to reconcile themselves with those they have wronged." (2001, 116) Some critics argue that expressing public condemnation is not a sufficient account of punishment, since punishment necessarily involves hard treatment, and resentment can be expressed without it. Nathan Hanna, for example, argues that we might express our condemnation by requiring payment of court costs or compensation (Hanna 2008, 137) or by confining people in a way comparable to involuntary psychiatric treatment, and thus without intending to cause

---

something that we "read off of" the action regardless of the psychological peculiarities of a wrongdoer's psychology that led him to commit the wrong." (Hampton 1992, 1684)

suffering (Hanna 2009, 242).[4] Call this the Enactment Problem: Why does the condemnation have to be enacted by hard treatment rather than merely symbolically expressed?

Putting the Enactment Problem aside for now, the expressive account of punishment, as such, leaves open the question of whether and how punishment is *justified*. The traditional justifications for punishment appeal either to the intrinsic value of the perpetrator's suffering and the perpetrator's deserving pain (retributivism) or to the benefits of hard treatment, such as the good social consequences of deterrence or incapacitation (consequentialism). An expressivist understanding of crime and punishment need not appeal to a new kind of justification altogether, but may instead provide a different insight into existing models (Kahan 1996, 601).

Jean Hampton's take on retributivism is a good example of this. For her, crime sends a message, and punishment counters or contradicts it with its own message. Hampton's case for this draws on the Wrong Valuation account of wrongdoing. The criminal's action makes a false moral claim and denies moral reality. Retributive punishment is "the defeat of the wrongdoer at the hands of the victim … that symbolizes the correct relative value of wrongdoer and victim" (Murphy & Hampton 1988, 125). It need not involve the infliction of pain (though that's often part of it), but rather mastering the would-be master himself in order to "deny the wrongdoer's false claim to superiority and to assert the victim's equal value" (ibid., 126). Thus, punishment can "annul the message" (ibid., 131) sent by the crime. Similarly, Kahan says, "The proper retributive punishment is the one that appropriately expresses condemnation and reaffirms the values that the wrongdoer denies." (1996, 602)

---

[4] One argument, put forward by Christopher Bennett, is that reactive attitudes as such involve "commitment to retribution, to the thought that it is non-contingently a good thing that those who have done wrong should undergo certain forms of suffering" (Bennett 2002, 147). For Bennett, this suffering that reactive attitudes seek is of two main varieties: the pain of social isolation, and the pain of guilt. Hanna denies this analysis of reactive attitudes. My sympathies are on Bennett's side, but I lack the space for proper discussion here.

I believe Hampton's view is on the right track. But it places too much weight on communicating a message about the relative value of the criminal and the victim, and the "evidence" that such a message supposedly provides (Hampton 1992, 1676). There is more than mere communication at stake, or so I will argue. Further, instead of Hampton's Wrong Valuation account of wrongdoing, I adopt the more general Expressive Wrongdoing view: wronging others is a matter of enacting inappropriate or unfitting attitudes towards them (where inappropriateness need not be understood in terms of value).

Punishment and Inviolability

To find an expressive rationale for punishment, we need not just look at what punishment says about the perpetrator, but also what it says about the *victim*. After all, punishment also sends a message to the victim: we, the community, will not let others enact disrespect towards you with impunity; we'll stand up for you. This is also suggested, but not fully developed, by Hampton, who notes that "whatever one's theory of human worth is, I am suggesting that societal punishment practices should be seen as created and designed to protect it" (Murphy & Hampton 1988, 141).

What is at issue is a special kind of value, which I'll call *inviolability*. Inviolability is a *deontic status* that something may have. To be inviolable is to be such that there are certain things others are not permitted to do to you. I will take it for granted that human beings are by default *de jure* inviolable: it is morally wrong to do certain things to them. For example, other things being equal, it is wrong to hurt someone against her will. If someone is treated in an impermissible way, then, other things being equal, it is *de jure* fitting to impose some kind of negative *sanction* on the agent. This is a conceptual truth: what it *is* to have a certain status is for sanctions for certain violations to be appropriate (and conversely, for sanctions for certain behaviours to be inappropriate – part of what it is to have the status of being the

owner of something is for it to be inappropriate for others to interfere with your using it). Normative status and fittingness (or unfittingness) of sanctions are simply two sides of the same coin. Calling something impermissible and then saying that no sanction is fitting amounts to contradicting oneself, unless there is an excuse or exemption for the agent.

Our inviolability confers us a kind of dignity or worth that is good for us independently of our actually being harmed. This is something that Thomas Nagel emphasizes when he says that "not only is it an evil for a person to be harmed in certain ways, but for it to be *permissible* to harm the person in those ways is an additional and independent evil." (Nagel 2002, 38) In a similar spirit, Frances Kamm (2005) talks about the importance of our inviolability over and above the importance of not being actually violated.

Moral inviolability is something that we can't lose as a result of the action of others, but we can meaningfully speak of a kind of *de facto* inviolability that we may or may not possess. What strips us of our *de facto* inviolability is *not* merely that someone harms us or wrongs or disrespects us. But if others can disrespect me *with impunity* or with a mere slap on the wrist, I will have little dignity left. Our *de facto* inviolability is thus constituted not by the fact that others treat us as rational and autonomous agents or fail to do so, but by the fact that others *must*, on pain of external sanction, treat us as such. This is the sense in which law can confer on us a social standing that befits an equal member of a democratic community. It can't stop others from engaging in a disrespectful behaviour such as stealing or assault, but it can penalize such behaviour and thereby manifest respect for everyone. If, conversely, the law *didn't* penalize such behaviour against a particular group, leaving them open to attack with impunity, that would amount to denying them equal respect and thereby constituting them as having a lower standing than others, *even if* no one ever actually attacked them.

So victim-focused LE says that enacting retributive attitudes is justified, because it is the only way to maintain or establish the victim's *de facto* status as inviolable – to guarantee that the victim *de facto* has the inviolability she already has *de jure*. Merely symbolically expressing censure of the perpetrator or respect for the victim does not suffice to establish the status. I emphasize that this is not a psychological but, as it were, a logical matter: to have a certain status is for it to be the case that behaviours inconsistent with it are sanctioned. To be inviolable is to be someone towards whom no one can violate without punishment – if someone enacts disrespect toward you, the state will enact condemnation of the perpetrator. Inviolability is a matter of degree, and as Hampton points out in her parallel discussion (1988, 141), historically certain groups of people – rich, white, male – have enjoyed higher de facto inviolability, since punishment for disrespecting them has been surer and harder. From this perspective, the Enactment Problem turns on a normative question: whether hard treatment is a proper expression of societal reactive attitudes depends on how inviolable the victims should be treated as, since inflicting suffering on a perpetrator is a way of expressing respect for the victim.

Let us turn to potential challenges to this account. First, Matthew Adler considers something like victim-focused LE in his discussion of Hampton. On his reading, Hampton says that wrongdoing "has led the community to believe that the victim has lesser moral worth" (2000, 1424), and punishment is necessary to reverse this status harm. He then has an easy time showing that punishment may not be the best or only way to change community norms and beliefs. But this misses the point altogether. The problem isn't that wrongdoing results in false *beliefs* about the victim's status, but that enacting disrespectful attitudes, if not countered, actually *changes* the (*de facto*) status itself.

Second, victim-focused justification of expressive punishment also avoids what I'll call the Illiberality Challenge. This challenge is that if punishment condemns the moral quality of the perpetrator's attitudes-expressed-in-action, it oversteps the boundaries of political liberalism, which requires moral neutrality. Bennett, for example, justifies state denunciation by appeal to "a model of the responsible or virtuous citizen, a model that represents an individual with concerns and attitudes consonant with the defining authoritative values of the polity", and says that "criminal law therefore enjoins people to live up to this model" (2006, 301). Though Bennett believes this is consistent with Kantian liberalism, the perfectionist worries of critics like Hurd (2001) are understandable. The victim-focused view, in contrast, does not need to appeal to moral improvement of the perpetrator as justification. The state must condemn crime in order to enact respect for the victim as an inviolable member of the community. There could hardly be a more liberal rationale for punishment.

4. Responding to Hate

In the previous section, I defended an Expressive Wrongdoing account of crime and a victim-focused Legal Expressivist account of the nature and justification of punishment in general. I will now argue that it provides the most promising framework for thinking about hate crime legislation.

On the Expressive Wrongdoing view, what makes something a crime is that it enacts attitudes that are incompatible with the legal status the victim enjoys as a citizen or occupant of a particular role. Unless society in turn expresses its rejection of such attitudes, the victim's status really is lowered: she is someone to whom it is (legally) acceptable to do such things. On victim-focused LE, this is what justifies punishment, in proportion to the victim's

status and the seriousness of the violation. Ordinary crime, motivated typically by perception of self-interest or a personal grudge, poses a threat to the victim's inviolability only as a *side effect*, however. The victim's status is no part of the criminal's concern.[5]

Hate crime is different. The attitudes it enacts – contempt, hate, disgust – contain an evaluation of the victim, by virtue of who she is (and hence everyone else who shares the relevant characteristic), as inferior. Contempt, for example, presents its object as ranking low in worth as a person, by the lights of some ideal the contemner (perhaps wrongly) endorses, in Michelle Mason's apt formulation (Mason 2003, 241).[6] In that sense, at least, these attitudes aim at lowering the status of the victim and those like her. And if those who are in positions of authority in the society do not respond to this appropriately, they do succeed. It does not suffice to reject the disrespect towards the victim that all crimes manifest to cancel out the hate and contempt. Indeed, failing to respond to this aspect of the crime leaves other members of the society complicit with it. So my main argument for hate crime legislation goes as follows:

1. Committing a base crime against someone as an interchangeable member of a group regardless of personal profit or pre-existing personal relationship manifests an attitude of disgust, contempt, hostility, or hate, for short rejection as an equal member of a democratic community, toward all members of the victim's group.

2. A criminal justice system that fails to impose an additional sanction for a crime that manifests rejection as an equal itself manifests indifference on the part of the public toward the rejection.

3. Social or legal status is in part constituted by dispositions to sanction manifestations of attitudes.

4. So, manifesting public indifference to rejection as an equal constitutes the status of all members of the victim's group as inferior, making the public complicit in the hate aspect of the crime.

---

[5] As a reviewer for this journal pointed out, this is not true of all non-hate crimes. I accept that my argument generalizes to some non-hate crimes, but lack the space to explore the implications here.
[6] Mason believes that contempt can be morally justified, if the ideal of the person is justified, and if the object really does fall short of it. I take it as a given that these conditions are not met in the case of hate crime.

5. Having an inferior status is in itself bad for a person, regardless of other physical or psychological harm.

6. So, equal respect for persons requires an additional sanction for the hate element of a crime.

The first premise delineates some of the grounds for concluding that a crime expresses a demeaning or degrading attitude toward a group *in addition to* whatever disrespect or disregard the base crime already manifests. If you burn my house because it's in the way of a new development you want to build, you disrespect me; if you burn my house because you don't want my kind to live in the neighbourhood, you also show contempt and disgust toward me and my people. The second premise points to the link between imposing sanctions and the attitudes of the public this manifests. The third one links sanctions to status, as discussed above. The fourth premise draws the preliminary conclusion that the public's attitudes manifested by the failure to sanction the hate aspect constitute the victim's group as inferior to others. This is a strong claim, of course. But it seems to parallel the case with disrespect: again, if we do nothing (or react only verbally) while someone is being assaulted, for example, or adopt norms that fail to direct us to do anything, we are thereby making the victim into someone *de facto* insignificant. If we do nothing to react to the hate or contempt aspect manifested by the hate crime, we acquiesce with the treatment of the victim as less significant than others, thereby lowering her *de facto* status.

The fifth premise says that an inferior status is bad in itself. In the real world, having such a status and being aware of it is of course also psychologically harmful. If you realize that no one is going to do anything if you are raped or beaten, this will have effects beyond the rape and beating itself. But as Nagel and Kamm pointed out, lacking inviolability is bad in itself. To put the point differently, it amounts to being dominated by another. As Philip Pettit (1997) has famously argued, even a slave who is never actually interfered with is

worse off than a free person, since the owner is always in a position to interfere with the slave's life on arbitrary grounds.

The conclusion, then, is that equal respect for persons, the kind of attitude that any legitimate liberal state is required to manifest toward all citizens, requires that crimes that embody hate or prejudice are punished more seriously than parallel crimes that do not embody the same attitudes. Contrary to a claim popular with critics, this is not punishment for *having* an attitude, but for *enacting* it, and furthermore enacting it in a way that is independently disrespectful of others in a way that merits state intervention (and thus should be criminalized in any case). Again, wrongdoing doesn't consist merely in harming others, but doing so in a way that embodies a devaluing attitude. Laws such as the Wisconsin Hate Crime Statute (Wis. Stat. § 939.645) that prescribe increased penalties for intentional selection of the victim on the basis of perception of protected characteristics seem to target the right thing.

5. Conclusion

In this paper, I have argued in defence of the view that both crime and punishment should be understood in expressive terms in general. Part of this defence was distinguishing the different ways in which attitudes can be expressed, symbolic expression and enactment. I have argued that although attitudes of hate or bias are in themselves beyond the rightful scope of legislation, those who manifest them by way of performing actions that are independently wrong and are reasonably interpreted as aiming to lower the status of their victims are morally liable to suffer enhanced punishment in order to protect the equal status

of the victim's group.[7] In order for us all to live in a society of equals, we must express our rejection of the rejection of equality that hate crime enacts by more than just words.

References

Adler, M. (2000). Expressive theories of law: A sceptical overview. *University of Pennsylvania Law Review*, 148, 1363–1502.

Al-Hakim, M. & Dimock, S. (2012). Hate as an aggravating factor in sentencing. *New Criminal Law Review*, 15 (4), 572–611.

Anderson, E. (1993). *Value in ethics and economics*. Cambridge, MA: Harvard University Press.

Anderson, E. & Pildes, R. (2000). Expressive theories of law: A general restatement. *University of Pennsylvania Law Review*, 148 (5), 1503–1575.

Bennett, C. (2002). The varieties of retributive experience. *Philosophical Quarterly*, 52, 145–163.

Bennett, C. (2006). State denunciation of crime. *Journal of Moral Philosophy*, 3 (3), 288–304.

Blackburn, S. (2010). Group minds and expressive harm. In *Practical tortoise raising*. Oxford: Oxford University Press, 64-89.

Brudholm, T. (2010). Hatred as an attitude. *Philosophical Papers*, 39 (3), 289–313.

Duff, A. (2001). *Punishment, communication, and community*. Oxford: Oxford University Press.

Feinberg, J. (1965). The expressive function of punishment. *The Monist*, 49 (3), 397–423.

Hampton, J. (1992). Correcting harms versus righting wrongs: The goal of retribution. *UCLA Law Review*, 39, 1659–1702.

Hanna, N. (2008). Say what? A critique of expressive retributivism. *Law and Philosophy*, 27 (2), 123–150.

Hanna, N. (2009). The passions of punishment. *Pacific Philosophical Quarterly*, 90 (2), 232–250.

[7] I have not said anything about symbolic expression of hate. I do not believe that the argument I have made can be extended to support hate *speech* legislation. It seems to me that the fact that others are free to say unpleasant things about my group does not undermine my status as inviolable in the way that selectively targeting me without punishment does, as long as I am similarly free to talk back or turn my back to it.

Hurd, H. M. (2001). Why liberals should hate “hate crime legislation”. *Law and Philosophy*, 20, 215–232.

Hurd, H. M. & Moore, M. (2004). Punishing hatred and prejudice. *Stanford Law Review*, 56 (5), 1081–1146.

Iganski, P. (2001). Hate crimes hurt more. *American Behavioral Scientist*, 45 (4), 626–638.

Kahan, D.M. (1996). What do alternative sanctions mean? *University of Chicago Law Review*, 63 (2), 591–653.

Kahan, D. M. (2001). Two liberal fallacies in the hate crime debate. *Law and Philosophy* 20, 175–193.

Kahan, D. M. & Nussbaum, M. (1996). Two conceptions of emotion in criminal law. *Columbia Law Review*, 96 (2), 269–374.

Kamm, F. (2005). *Intricate ethics*. New York: Oxford University Press.

Lawrence, F. (1999). *Punishing hate*. Cambridge, MA: Harvard University Press.

List, C. & Pettit, P. (2011), *Group agency*. Oxford: Oxford University Press.

Mason, M. (2003). Contempt as a moral attitude. *Ethics*, 113 (2), 234–272.

Nagel, T. (2002). *Concealment and exposure and other essays*. New York: Oxford University Press.

Pettit, P. (1997). *Republicanism: A theory of freedom and government*. Oxford: Clarendon Press.

Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25.

Tuomela, R. (2007). *The philosophy of sociality*. New York: Oxford University Press.