



This is a repository copy of *Retributivist Theories*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/95450/>

Version: Accepted Version

Book Section:

Bennett, C.D. (2013) *Retributivist Theories*. In: Bruinsma, G. and Weisburg, D., (eds.) *Encyclopedia of Criminology and Criminal Justice*. Springer , New York , pp. 4446-4456. ISBN 9781461456896

https://doi.org/10.1007/978-1-4614-5690-2_603

© 2013 Springer. This is an author produced version of a chapter subsequently published in the 'Encyclopedia of Criminology and Criminal Justice', available at Springer via <http://link.springer.com/referencework/10.1007/978-1-4614-5690-2> . Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

TITLE:**Retributivism****BYLINE:**

Christopher Bennett
Department of Philosophy
University of Sheffield
U.K.
c.bennett@sheffield.ac.uk

SYNONYMS:

Retribution
Retributive Punishment

OVERVIEW:

This essay looks at recent retributivist theories that draw on denunciation and the expression of moral emotions in order to justify punishment. After setting out some of the canonical sources of the retributivist tradition, and explaining some of the most serious objections to this tradition, it looks at how these recent developments seek to overcome the objections while preserving what seems most of value in retributive ideas. The essay identifies work by P. F. Strawson and Patrick Devlin as the starting-point of these developments. In different ways Strawson and Devlin seek to vindicate the idea that punishment should express our sense of the moral seriousness of crime as wrongdoing. Furthermore, they imply that punishment is necessary to do justice to the moral seriousness of wrongdoing. The essay considers to what extent this line of argument, if more fully developed, might offer a successful defence of retributivism. It also gives a survey of some of the major recent contributions to this field, including Jeffrie Murphy, Andrew von Hirsch and R. A. Duff.

MAIN BODY TEXT:**1. Introduction**

Three important developments in recent thinking about the justification for punishment are: the rise of communicative theories of punishment; renewed interest in the moral emotions; and the return of retribution. This essay concerns the way the first two have informed the third. After a long period of neglect, during which it was thought to be a barbaric remnant of superstitious thought that more rational forms of society would sweep away, retribution has, over the past forty years, returned to the fore as a rationale for punishment. There are, of course, various social and political factors that have contributed to this change. But a full explanation of the trend would surely have to refer to the fact that theorists over the past forty years have discovered new and prima facie attractive ways of construing retributive ideas. One claim, associated with early work by Herbert Morris (Morris 1968) and Jeffrie Murphy (Murphy 1973) amongst others, is that retributive punishment is necessary to restore the fair balance of benefits and burdens in a system of social cooperation. However, this approach has always faced the charge that it reduces all retributive action to action aimed at rectifying unfair advantage, and therefore cannot capture the sense that there is something about wrongs such as murder, rape assault and wilful neglect that call for retributive responses. A different strand of the new thinking about retributivism, on the other hand, has concentrated on the connections between retribution, moral condemnation

and the moral emotions. Wrongdoing is something that we often feel strongly about – and such feelings can, it is said, (sometimes at least) be justifiable. It is this latter strand of retributivism that is the subject of this essay.

2. What is retributivism?

Whether there is something yet living in the retributive tradition depends on whether, on due reflection, at least some of its canonical ideas continue to strike us, under a certain interpretation, as moral insight rather than illusion.

One central source for this tradition is the idea, common to Judaism, Christianity and Islam, that there is one perfectly just God who will, at some ultimate point of judgement, reward the virtuous with perfect happiness and punish the vicious with eternal suffering. Immanuel Kant captures this thought by claiming that the highest conceivable or possible good is the “distribution of happiness in exact proportion to morality [i.e. to the goodness of one’s will]” (Kant 1996: 135). On this view, the purpose of retributive action is to help bring about a perfectly just state of affairs. Such action may be seen, in the final analysis, as the business of the Deity. However, it is important to recognise that this is not the only way to think about retribution.

Another source of the retributive tradition is deontic or act-focused, and thinks of justice as requiring that transgression of authoritative limits on conduct must be answered by a like, or at any rate commensurate, infringement inflicted on the transgressor. For instance the claim, from the Book of Exodus, that a wrongdoer should give “life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound, for wound, stripe for stripe” (often called the *lex talionis*) can be interpreted in this light. A similar idea can be found in Pre-Socratic Ancient Greek thought (for some discussion, see Nussbaum 1993).

Thirdly, the retributive tradition contains ideas about moral contamination or pollution. For instance, we might think of the forbidding claim made by Immanuel Kant that, were a civil society to disband, it would be necessary for the last murderer in prison to be executed before the disbandment should take place, to prevent “blood-guilt” from infecting each of the departing members (Kant 1991: 142). The blood-guilt presumably affects the murderer most directly. But, secondly, the contamination can be passed on to otherwise innocent parties if they do not do their duty in visiting proper justice on the murderer. Again, a similar conception is present in Ancient Greek thought, where the failure of Thebes to take action against their king Oedipus has dreadful consequences for all.

Fourthly, the retributivist can lay claim to the ideas of repentance and atonement common again to a number of religious traditions (Etzioni and Carney 1997). The idea of at-one-ment suggests a moral fracture in the offender, a separation from God or the Good, that is brought about by sin or wrongdoing. The wrongdoer needs to be made whole again, to overcome her separation from what is most important in her life (i.e. God or the Good) through sincere repentance and atonement. One prominent development of this set of ideas is the Roman Catholic sacrament of penance, in which penitents undertake something otherwise harmful or onerous in order to make amends for sin.

What each of these sources has in common is the view that it is inherently right or good that something negative should be done to those who do wrong. One feature commonly said to be distinctive of retributivism is that it is a backward-looking theory: it takes the importance of the wrong committed to be the central reason for punitive action. We can understand better what is meant by this if we contrast it with forward-looking approaches such as deterrence. According to a deterrent theory, the reason we punish is, not the moral nature of the crime itself, but rather the need to prevent future crimes: punishment is necessary because (and insofar as) it produces some future good. The forward-looking view gives a clear explanation of why punishment is important, but also makes its justification derivative and conditional. Whether punishment is justified depends on whether it really is an effective means for the production of the future good we have in mind. On the backward-looking view, however, there is something about punishment that is inherently fitting for answering, or annulling, or avenging, etc., the wrong. Thus an aspect of the retributive view is that some acts derive their importance and necessity, not from the future benefits that they bring about, but rather from their role in doing justice to what has gone before.

In addition to its backward-looking focus, another distinctive feature of retributivism is that it places some importance on harm suffered by the offender as a result of the wrongdoing. Thus, according to one prominent recent characterisation, retributivism centres on “the Desert Thesis”: “that when a person has done something that is morally wrong it is morally better that he or she should suffer some loss in consequence” (Scanlon 1999: 274). If we put these two features of retributivism together we come to the distinctive retributivist claim that it is morally important or necessary that wrongdoing or evil should lead to some response that involves the offender suffering in some way.

3. Serious objections to retributivism

The shape of any contemporary retributivist theory will be dictated by the need to preserve what seems most attractive in the retributive tradition, while at the same time avoiding or answering the main objections that have been raised to retributive action. We turn to look at some of these main objections now.

First of all, critics often object to retribution as a rationale on the grounds that it doesn't bring about any good. This is connected with its backward-looking orientation. It might be claimed that retributive action doesn't actually achieve anything. Indeed, the situation is worse than this, because retribution involves glorying in suffering that seems to be inflicted for its own sake. Therefore retribution can look redundant at best, and at worst, downright sadistic.

In response to this, however, the retributivist should deny that nothing is achieved by retribution (Murphy 1971). The retributivist thinks that sometimes it is necessary to take action to do justice to a past act, independently of future benefit. Actions of gratitude, or reward, or mourning seem to share this feature. What such actions achieve is an appropriate response to that past action. Therefore the retributivist should resist the claim that retributive action is redundant. And if the suffering of the offender is part of such appropriate response, satisfaction at suffering is not mere sadism.

Nevertheless, there is clearly an important question to be answered as to what makes the suffering of the offender the appropriate response to wrongdoing. The suspicion will be that the retributivist is influenced by ideas about moral pollution. But those ideas might be said to be part of a world view that we have long outgrown (see e.g. Smart in Smart and Williams 1973 on guilt as a stain on the soul).

A second major line of criticism looks at the problem of free will. For retributive responses to be appropriate, it would have to be the case that perpetrators of wrongdoing acted freely. If the apparent wrongdoing was not really the wrongdoer's fault then they cannot deserve to suffer punishment in response. But, the objection goes, our understanding of the basis of human action is now sufficiently advanced that we can see that no action is entirely freely chosen. Human beings are complex biological mechanisms; they act in social conditions they have not chosen, on the basis of culturally constructed understandings. Where is the room for free will in all this?

We will look in more detail at responses to this criticism below. However, it is worth saying at this point that empirical evidence about the basis of human action will not by itself refute moral responsibility. In order to assess the justice of this criticism of retributivism, we would have to look first of all at which forms of freedom are necessary for moral responsibility. This is not an empirical question, but a philosophical and moral question about the conditions under which it is fair and right to hold someone to account for what she has done. The absence of some sorts of freedom does indeed excuse or exempt from blame. But which? Only once we have answered that question can we turn to the empirical evidence and look at whether, given everything we know about human beings, it is plausible to think that we are free in the ways moral responsibility requires. Therefore an assessment of the free will criticism depends on some understanding of what is going on in retributive action, and what forms of unfreedom make retributive action invalid.

Some other common criticisms of retributivism are worth mentioning at the moment. For instance, if we are committed to retribution, to what extent should we take into account the extent to which someone has already suffered in their life in assessing how much they presently deserve? More broadly, and leading on from this, doesn't the goal of apportioning happiness to personal virtue (and suffering to personal evil) require a degree of knowledge of people's character and motivation that we routinely lack (perhaps it is no accident that the task of doing justice is often allotted to an all-knowing God). One way to respond to these problems is for the retributivist to reject a character-based concern with personal virtue and vice as a whole (the first source considered above), and argue instead for an act-focused form of retribution as a response to transgression (the second source). However, even the act-focused form of retributivism will have to address the question of how much suffering is proportionate to which crime. Once we reject the notion of literal equivalence between punishment and crime – for instance, on the grounds that it would have us raping rapists or torturing torturers – the notion of proportionality becomes more mysterious.

Another type of criticism asks whether retribution is an important, or even legitimate, state purpose. This is compatible with allowing that some form of retributivism may be morally important. But if we think of how much money is spent on criminal justice, we have to ask whether retribution gives good value for that money. If prison is necessary to bring about

security of person and property, it may be money well spent. But why is it important to spend the money on something backward-looking? A related criticism is that if retributivism were the rationale for state punishment then the state would be committed to punishing for all moral wrongdoing. So for instance, lying, adultery, sexual perversions ... if any of these are moral wrongs then, the objection claims, the retributive state would punish for them. Yet this would be illiberal: it is characteristic of the value given to freedom by the liberal state that it at least some moral wrongs are viewed as private and irrelevant to the legitimate concerns of law. In short, retributivism seems allied to legal moralism – the claim that law should sanction immorality as such – whereas the liberal position, it might be said, rejects this approach.

4. Retribution, denunciation and emotion

Having looked at some of the sources of the retributive tradition, and some of its most serious objections, we can now turn to recent developments that seek to show how retributivism might yet have the resources to answer these concerns. It is central to the approach to retributivism that we will look at here that retribution is not only a rationale for punishment, but also a feature of our emotional life in a range of interpersonal relationships. Thus, according to this line of thought, there is a deep connection between retributivism in punishment and our intuitive ways of dealing with one another in everyday life. In one way this represents a strength of retributivism over e.g. the view that punishment is justified as deterrence or incapacitation. After all, one of the reasons why people maintain and support an institution of punishment might well be that it reflects something important about the way they think wrongdoers should be treated in non-legal cases, for instance, when one person betrays or lies to or steals from or unjustifiably harms another. On the other hand, retributivists may be vulnerable to the criticism that they fail to recognise the extent to which legal punishment in a diverse modern society differs from blame in intimate relationships.

We can begin by identifying two interventions that helped to initiate the current revival in thinking about retributivism, condemnation and the emotions: P. F. Strawson's "Freedom and Resentment" (Strawson 1962); and Patrick Devlin's "The Enforcement of Morals" (Devlin 1959). The latter initiated a debate with H.L.A. Hart (Hart 1963), one result of which was the publication of Joel Feinberg's highly influential "The Expressive Function of Punishment" (1965 – see also Moberly 1968). The argument of neither, of course, has been universally accepted. Yet both operate by identifying something missing in the then-prevailing liberal-utilitarian consensus on criminal justice, with its emphasis on the efficient management of social behaviour; and both suggest that the "barbaric" retributivism opposed by utilitarianism has more to be said in its favour than previously thought.

Devlin argued for retaining the criminalisation of homosexuality in the UK on the basis that it was quite proper for the criminal law to target immorality or sin that arouses repugnance in "the reasonable man." Now there are many points in the Hart-Devlin debate in which Devlin the reasonable reader may find Devlin to be in the wrong. For instance, many people will find his views on homosexuality repellent; and his view that criminal law should be explicitly recognised as having a basis in Christianity is hard to reconcile with the multicultural nature of many modern democracies. Furthermore, it is often unclear whether he is arguing that criminal law should be based in what is justifiably regarded as immorality, or rather what the "common man" finds repugnant. We may share Hart's view, in *Law, Liberty and Morality*, that the criminal law

ought to be shaped by the moral ideas that best survive critical reflection, rather than our gut feelings. Nevertheless, when Devlin argues that “no society can do without intolerance, indignation and disgust; they are the forces behind the moral law” – and makes it clear that he is talking about disgust “that is deeply felt and not manufactured”, the presence of which “is a good indication that the bounds of toleration have been reached” – he touches upon an important aspect of the function of criminal law that is ignored if we see punishment merely as a technique for behaviour control. This is a point picked up by Feinberg in his discussion of the expressive aspect of punishment. Punishment is not merely, as the utilitarian might regard it, an incentive against anti-social behaviour; rather it is a vehicle of symbolic collective condemnation, expressing the anger, indignation or repugnance that an act arouses in the “right-thinking person.” What is missing in purely forward-looking accounts of the criminal sanction is something that connects punishment with our sense of the seriousness of the wrongdoing.

Strawson’s argument, on the other hand, is addressed most directly to the debate over free will and moral responsibility. As I explained above, any argument about whether we have free will depends on some conception of the kind of free will necessary for moral responsibility. Strawson gives us an account of what we are doing when we hold people morally responsible, from which he then adduces an account of the freedom we need to be morally responsible. He argues that our sense of moral responsibility and retributive justice is found in the operation of moral emotions such as resentment and indignation, which lead us to withdraw goodwill from, and inflict harm on, those to whom they are directed. Strawson defends retribution and responsibility by arguing that the tendency to have such emotions is inextricably intertwined with accountability in interpersonal relationships. The interpersonal relationships that Strawson has in mind are those grounded in mutual expectations of some level of goodwill: relationships that are based in trust rather than mere containment and accommodation. Such relationships are an important part of the value of human life, and do not seem to be a merely illusory ideal. But when we enter into such relationships we adopt a certain practical perspective on human behaviour according to which participants are reasonably subject to certain demands for good will. Of course, this perspective on human action is not the only one that it is theoretically possible to adopt. We can also see human action, not as a response, adequate or inadequate, to legitimate expectations, but rather as the deterministic behavioural output of a number of personal, subpersonal and environmental inputs. This is what we do when we note that human beings are biological mechanisms, that their action is the result of social forces, etc.. It is also the perspective we adopt when dealing with the very young, or the insane. But human life would be barren – perhaps even impossible – if this perspective (which Strawson calls the objective attitude) were the only one we ever took up. Therefore we should reject the abandonment of moral responsibility; there is no reason to think that our practices of accountability are in bad order.

Strawson’s arguments, and the outcome of the Devlin-Hart-Feinberg debate, altered the way in which it was possible to understand the significance of retribution. We can see this if we think about how the viewpoint they developed might be used to help us address some of the criticisms of retribution we considered above. One of the main criticisms of retribution was that its backward-looking nature makes retributive action redundant, and its desire for the suffering of the offender makes it cruel. Another problem concerns free will; while another argues that apportioning virtue to happiness, even if desirable, is a pointless task for limited human beings. Yet another criticism is that retributivism is tied to outdated ideas of moral pollution. The

Devlin-Strawson approach, in response, sees punishment or blame in deontic fashion, as a necessary part of recognising the significance of responsible wrongdoing in the context of relationships that are based on trust. Being subject to these retributive reactions for wrongdoing is simply a condition of being included in such relationships.

Strawson and Devlin enable the retributivist to argue that the value of retributive responses lies in the fact that they a) recognise the seriousness of the wrongdoing, and b) recognise the agent who did the wrong as a human being who could and should have done better, and can thus justly be held to the relevant moral demands. When a person is punished, in other words, it involves being treated harshly, but it also involves being included rather than excluded from the terms of the relationship.

This line of thinking might also allow the retributivist to turn the tables on the deterrence theorist, charging that the deterrence theorist advocates merely threatening the offender, treating her as a mere means to a social end, whereas retribution is directed at the offender as an individual, a responsible moral agent. This criticism is made most forcefully by R. A. Duff, who quotes approvingly Hegel's damning remark that, according to the deterrence theory, punishment is like raising a stick to a dog (Duff 1986: p. 180). Even if this seems too strong, a more modest point can be made. Strawson reminds us that mere deterrence is not what we are engaged in when we experience emotional correlates of punishment like resentment and indignation. This raises the question whether punishment itself could not be made more like an interpersonal interaction. Deterrence theory sees punishment as an interaction between an individual and a bureaucratic state agency motivated by an overview of social welfare. A potential strength of the retributive approach is that it envisages the criminal justice process as - ideally - something more like an interpersonal interaction.

5. Two cheers for Devlin and Strawson?

In order to understand more recent developments, we turn now to evaluate how far these arguments take us. They may have convinced us that offenders should be held to account for what they have done, and that their actions deserve criticism or moral disapproval. Furthermore, we might be persuaded that condemnation of at least a certain range of wrongful actions is a legitimate state purpose. Perhaps the state has no business condemning people for adultery, or for sexually perverted behaviour. But we might agree that it should condemn those who, say, violate the legal rights of others, or who rape, murder and assault. This view of punishment provides an important alternative to the view that the function of criminal justice is simply as an effective means to protect life, liberty and property, and has been explored by writers such as Andrew von Hirsch (von Hirsch 1985, 1993) and R. A. Duff (Duff 1986, 2001)

However, the argument so far seems to suffer from an important gap. Serious wrongdoers may deserve condemnation, perhaps even by the state. But why, we might ask, do offenders deserve to suffer in addition to being criticised? Feinberg himself advances this criticism (Feinberg 1965). He argues that punishment is merely a conventional symbol of moral disapproval, and asks whether we could not find less harmful symbols to do that work for us. This question requires an answer, and, in response, censure theorists find themselves in two camps: those who argue that there can be no meaningful condemnation without hard treatment; and those who agree with Feinberg that the goal of censure does not by itself require hard treatment. In the latter

camp lies the influential version of the censure conception of punishment put forward by von Hirsch. Von Hirsch argues that censure is an important function of punishment because it addresses the crime as a wrongful action, does justice to the victim, and conveys to the offender that some response is required to the offence (von Hirsch 1993, pp. 9-10). However, he argues that the hard treatment element of punishment is only necessary for preventive purposes: if punishment had no deterrent effect there would be no need for hard treatment. A similar view is put forward by Matt Matravers (Matravers 2000). A prominent writer in the opposed camp is Igor Primoratz (Primoratz 1989; see also Moberly 1968), for whom punishment is analogous to language. If we change the symbols, according to Primoratz, we change what is said or done by punishment, so the symbolism of punishment is not conventional in the sense that we can arbitrarily choose what symbols to use without change in meaning. If we aim to condemn then we must use a vehicle that is symbolically adequate.

This raises a question about the nature of symbolic or expressive action in general, and as such takes us into deep waters beyond the philosophy of criminal law and punishment. To what extent can we alter the symbolic action we use to convey e.g. gratitude or grief or outrage while still preserving a meaningful connection to these attitudes? The retributivist – who wishes to argue for some infliction of hard treatment as the privileged symbolism for expressing outrage – requires some view on which symbols are more or less resistant to alteration. Thus in response to the objection that we might respond to wrongdoing by turning the other cheek, or simply communicating disapproval in a calm and measured way, the retributivist will want to argue that these responses, insofar as they avoid inflicting proportionate hard treatment, fail to do justice to the seriousness of the wrongdoing involved. One way to defend such a view, which can be seen in recent work by Jeffrie Murphy and Jean Hampton (Murphy and Hampton 1988), is to follow Strawson and Devlin in referring to the emotions. On the basis of phenomenological investigation, we can say that the experience of being in the grip of an emotion provides a subject with a deep sense of fit between some feature of her situation and the behaviour that expresses that emotion, such that that feature appears to “call for” that behaviour. For instance, if I have cause for pride I might find myself “walking tall”: and in that situation walking tall feels right. Or if I perceive myself as having done something shameful then covering my face or hiding away feels right. With respect to retribution, we might point to something to do with the fact that the emotions of outrage, indignation, resentment are expressed through aggressive action taken against the offender, such that it appears to us that the situation of wrongdoing calls for such a response. This might explain our sense that the symbolism of punishment is at least in part invariable: in order to express condemnation symbolically we must do something that connects with the behaviour we find satisfying as an expression of the emotions aroused by wrongdoing.

6. Morality, retribution and the emotions

Given the history of hostility to the emotions in Western philosophical thought, however, this turn to the emotions to inform our thinking about punishment is likely to provoke strong criticism. Thus a critic might ask whether the retributivist is right to give such weight to the emotions in normative theorising – theorising that may, after all, end up with some people (offenders) being seriously harmed. In particular, we might wonder whether, from the fact that something feels right when in the grip of an emotion, we can conclude anything about the nature of right or permissible action. This objection need not deny that there can be some value in the

expression of emotion. Perhaps in spontaneous interpersonal interactions we cannot help expressing our emotions. And perhaps we shouldn't always try to restrict such expression: maybe there is a value in spontaneity. But the spontaneous expression of retributive emotions leads to suffering; and the institutionalised expression of those emotions through state punishment can blight whole lives. The suspicion will be that our emotionally-driven sense of fitting response is too slender a thread on which to hang such weighty matters. Indeed, criminal justice policy, according to this challenge, may be better seen as containing rather than releasing such emotions (e.g. see Gardner 1998 on the "displacement function" of punishment).

In exploring how the retributivist might answer this criticism, we will need to say something about recent developments in the philosophy of emotion and moral epistemology. But briefly, the retributivist might attempt to argue three points. First of all, that emotions are not brute psychological states, but are what Scanlon calls "judgement-sensitive attitudes" (Scanlon 1999), attitudes in principle responsive to justificatory reasons. Secondly, the reason that emotions are judgement-sensitive attitudes is that they present a subject with an evaluatively-loaded (cognitive) representation of the situation under consideration, the plausibility of which can be assessed by the subject for its coherence with her other factual and evaluative beliefs. Thirdly, because the emotions can be understood as intelligent, cognitive evaluations of our environment, and not merely automatic responses to it, we have a default epistemic entitlement to trust the way they represent our situation, in the absence, that is, of strong reason to doubt that representation. If those three points could be established then the retributivist and the critic might agree on the following possibility: that the way a situation is presented to us in the grip of an emotion could be our best way of understanding the evaluative nature of that situation and the response it demands of us. The argument with respect to retributivism would then come down to whether we have strong reason to mistrust those experiences in which it appears to us that taking wrongdoing seriously requires some type of hard treatment of the offender.

The retributivist's line of argument might start by following Nussbaum and Kahan in distinguishing a mechanical from an evaluative conception of the emotions (Nussbaum and Kahan 1996). The mechanical view sees emotions as strong sensations, immune from thought and rationality, that are triggered by certain stimuli and drive us to certain actions. One of the key developments in the philosophy of emotion in the twentieth century has been the observation that, at least in some respects, and for what might be called the "higher emotions" (Griffiths 1997), this view is too simple. The emotion of fear is not a simple sensation like the feeling of stepping into a warm bath. Fear is a state with intentional content, it is directed towards something (i.e. it is "fear of" or "fear about" or "fear that") in a way that simple sensation is not. Therefore fear could be said to be a state that represents the world in a certain way. It is a way of seeing or understanding (according to Roberts 1988, "construing") a situation rather than just simple feeling triggered by the situation. (This is one of the main arguments against the James-Lange hypothesis that emotions are simply the effects in consciousness of bodily changes brought about by external stimuli.) Nussbaum and Kahan use this as evidence that emotion is more like a cognitive or evaluative state than the mechanical view allows (though for criticism, see e.g. Deigh 2008). We could also point to the fact that it is often possible to judge emotions according to rational standards of adequacy to the facts: if I am afraid of a noise that turns out to be only the wind rattling the door-knob then normally speaking my fear will abate; this suggests

that fear is at least sometimes sensitive to reasons for that fear, that it can be justified or unjustified. Hence the claim that emotions are “judgement-sensitive.”

Any proponent of the cognitive or evaluative view of the emotions needs to recognise, however, that there is a partial truth in the mechanical view: namely that some emotions in particular are capable of getting out of proportion quite easily, and that they can grip our attention to the extent that we find it difficult to get other things into proportion. But while the existence of irrational and disproportionate emotion is clear, this should not blind us to the fact that it is often possible to bring our emotions into line with our considered evaluative beliefs. In other words, some emotions can be justified in the sense that they embody our considered values. Furthermore, if this is allowed it may open up the possibility that sometimes the direction of revision goes the other way round, and that it is our evaluative beliefs that have to change to accommodate the epistemic deliverances of an emotional experience that appears authentic and compelling. This is a possibility famously defended by Bernard Williams (Smart and Williams 1973, Williams 1976): although Williams doesn't put it quite like this, we might say that it is the idea that justified and proportionate emotion might involve seeing the world in the right light (cf. McDowell 1985). If this is at least a possibility then we might argue that we have a default entitlement to trust our emotions (or moral intuitions) unless we have reason to think that they cannot be made coherent with our other justified evaluative or factual beliefs. (Though the proper role of intuition in moral theory is the subject of great controversy: see Sinnott-Armstrong, Young, and Cushman 2010). Applied to retribution, this would mean that we have a default entitlement to trust our feeling that the retributive response is fitting unless we have reason to think that this response cannot be made coherent with the demands of those other values in the validity of which we have great confidence (Moore 1988).

Nevertheless, the most we have established thus far is the possibility of justifying retribution on the basis that we might have an undefeated entitlement to trust the way our emotions present the situation of wrongdoing. In order to determine the success of this strategy for justifying retribution we would need to investigate whether the purported validity of retributive experience in informing us of the demands of the situation has to be regarded as illusion because it conflicts with the other things we know about what the situation demands. And the reader might be forgiven for being underwhelmed by this result. For it might appear that it simply returns us to the realm of normative argument between the retributivist and her opponents about whether retributive responses are compatible or incompatible with what we know about our obligations e.g. not to impose suffering. However, if the “retributive emotions” strategy is successful then this reaction would be unjustified for three reasons. First of all, the success of the strategy would make it clear that the burden of justification lies with the opponent of retributivism. Retributive reactions couldn't be written off as mere “gut reactions.” Secondly, it would provide a way of understanding ideas such as moral pollution or contamination that doesn't involve a commitment to unwanted metaphysical baggage. “Moral contamination” would simply be that quality that, when we are in the grip of (let us suppose) justified retributive emotion, seems to us to call for that reaction. It would not be a property of the universe with the power to cause famine, pestilence, etc.. Thirdly, it would affect our view of how to conduct moral inquiry. In order to evaluate retributive reactions, what we would need to look at is, in part, whether we can understand our experience of those reactions as inauthentic and illusory, or whether it is rather

our conflicting beliefs that need to change. This introduces an essentially phenomenological Gestalt-like element into moral inquiry.

The likelihood that we will conclude that retributive reactions are authentic or illusory no doubt depends on which emotions we focus on. The relevant emotions are often taken to be those felt by third parties towards offenders, such as in Jean Hampton and Jeffrie Murphy's discussions (Murphy and Hampton 1988). But many readers may find that it is easier to feel comfortable with the claim that offenders should feel bad towards themselves than it is with the claim that "we" should endorse aggressive feelings towards "them." For this reason another important line of work for retributivists is to be found in the conceptions of guilt and remorse put forward by Herbert Morris (Morris 1981), Gabriele Taylor (Taylor 1985), R. A. Duff (Duff 1986) and Raimond Gaita (Gaita 1991).

7. Emotion and Punishment

By way of conclusion, we can now distinguish three ways in which punishment might be related to the expression of the emotions.

- Firstly, state punishment might be carried out in order to displace aggressive retributive action motivated by strong feeling on the part of individual citizens.
- Secondly, punishment might be carried out in order to maintain the perception of the legitimacy of the state amongst the populace, a perception that might be strengthened if state action is carried out in line with the strong feelings of the citizens.

What we need to notice about these first two suggestions is that their motivation comes, not from the endorsement of the emotions involved, but rather from the desire for social stability: the policy is driven by the need to contain or accommodate strong emotion that might disrupt social relations. The justifiability or unjustifiability of those emotions is not relevant to the social need to contain them. However, the result of this is that a theory that gives this kind of role to the emotions is not distinctively retributive. For the need for punishment derives in the end from the need for social stability, not the retributive concern with doing justice.

For a distinctively retributive view, we need to consider a third possibility, which comes about if, after due reflection, we can endorse the way the situation of wrongdoing appears to us when we are in the grip of retributive emotions. If state punishment is a vehicle for treating the offender as those emotions prompt us to treat him then we might say the following:

- Fourthly, state punishment is carried out because the way of seeing the significance of wrongdoing that is involved in the retributive emotions - namely, that it calls for hard treatment in response - is our best understanding of what the situation demands.

The retributive emotions present wrongdoing as mattering in a certain way, specifically as being impermissible or transgressive; and they present the situation as one that calls for a certain kind of response directed at the wrongdoer. The strand of retributivism in which we are interested here argues that we have reason to allow this phenomenology to inform our normative judgements, broadly endorsing our tendency to act on retributive emotions, and deploying state

punishment as a vehicle for the condemnation of the offender. The reason for expressing the emotion is not a desire to release emotional pressure – as though the emotion were just a psychological force – but rather because retributive responses are our best understanding of what the situation demands.

RELATED ENTRIES:

CIVIL REMEDIES; DETERRENCE; INCAPACITATION; NATURE OF MORAL CRIMES; NEW PENOLOGY; PENAL ABOLITIONISM; PENAL JUSTICE AND SOCIAL INJUSTICE; PENAL PHILOSOPHY AND SENTENCING THEORY; PROPORTIONALITY; REHABILITATION; RESTITUTION; RESTORATIVE JUSTICE: THE IDEA AND ITS HISTORY; RETRIBUTIVISM: TRADITIONAL; SENTENCING GUIDELINES; THEORIES OF PUNISHMENT; WHEN PUNISHMENT IS REHABILITATION.

RECOMMENDED READING AND REFERENCES:

J. Deigh, *Emotions, Values and the Law* (Oxford: Oxford University Press, 2008)

P. Devlin, “The Enforcement of Morals,” *Proceedings of the British Academy* 45 (1959)

R. A. Duff, *Trials and Punishments* (Cambridge: Cambridge University Press, 1986)

R. A. Duff, *Punishment, Communication and Community* (Oxford: Oxford University Press, 2001)

A. Etzioni and D. Carney (eds), *Repentance: A Comparative Perspective* (Lanham: Rowman and Littlefield, 1997)

J. Feinberg, “The Expressive Function of Punishment,” *The Monist* 49, pp. 397–423

R. Gaita, *Good and Evil: An Absolute Conception* (London: Macmillan 1991)

J. Gardner, “Crime: In Proportion and in Perspective” in A. Ashworth and M. Wasik (eds), *Fundamentals of Sentencing Theory: Essays in Honour of Andrew von Hirsch* (Oxford: Oxford University Press, 1998)

P. Griffiths, *What the Emotions Really Are: The Problem of Psychological Categories* (Chicago: University of Chicago Press, 1997)

H. L. A. Hart, *Law, Liberty and Morality* (Stanford: Stanford University Press, 1963)

I. Kant, *Critique of Practical Reason*, trans. T. K. Abbott (Amherst: Prometheus Books, 1996)

I. Kant, *Metaphysics of Morals*, trans. M. Gregor (Cambridge: Cambridge University Press, 1991)

- J. McDowell, "Values and Secondary Qualities" in T. Honderich (ed.), *Morality and Objectivity: A Tribute to J. L. Mackie* (London: Routledge, 1985)
- M. Matravers, *Justice and Punishment: The Rationale of Coercion* (Oxford: Oxford University Press, 2000)
- W. Moberly, *The Ethics of Punishment* (London: Faber 1968)
- M. Moore, "The Moral Worth of Retribution" in F. Schoeman (ed.), *Character, Responsibility and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press, 1988)
- H. Morris, "Persons and Punishment," *The Monist* 52 (1968), pp. 475–501
- H. Morris, "A Paternalistic Theory of Punishment," *American Philosophical Quarterly* 18 (1981), pp. 263-271
- Murphy Jeffrie G., 1973, "Marxism and Retribution," *Philosophy and Public Affairs*, 2: 217–43
- J. G. Murphy, "Three Mistakes About Retributivism," *Analysis* 31 (1971), pp. 166-9
- J. G. Murphy and J. Hampton, *Forgiveness and Mercy* (Cambridge: Cambridge University Press, 1988)
- M. Nussbaum, "Equity and Mercy", *Philosophy and Public Affairs* 22 (1993), 83-125
- M. Nussbaum and D. Kahan, "Two Conceptions of Emotion in Criminal Law," *Columbia Law Review* 96 (1996), 269-374
- I. Primoratz, "Punishment as Language" *Philosophy* 64 (1989), pp. 187-205
- R. Roberts, "What an Emotion Is: A Sketch," *Philosophical Review* 97 (1988), pp. 183-209
- T. M. Scanlon, *What We Owe To Each Other* (Cambridge, Mass.: Belknap Press, 1999)
- W. Sinnott-Armstrong, L. Young, and F. Cushman, "Moral Intuitions" in J. Doris (ed.), *The Moral Psychology Handbook* (Oxford: Oxford University Press, 2010)
- P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962), pp. 1-25
- J. J. C. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973)
- G. Taylor, *Pride, Shame and Guilt* (Oxford: Clarendon Press, 1985)

B. Williams, 'Moral Luck', Proceedings of the Aristotelian Society 50 (supplementary volume) (1976), pp. 115-35

A. von Hirsch, Past or Future Crimes: Deservedness and Dangerousness in the Sentencing of Criminals (New Brunswick: Rutgers University Press, 1985)

A. von Hirsch, Censure and Sanctions (Cambridge: Cambridge University Press, 1993)