

Hume on Responsibility and Punishment

Author(s): Paul Russell

Source: *Canadian Journal of Philosophy*, Vol. 20, No. 4 (Dec., 1990), pp. 539-563

Published by: Canadian Journal of Philosophy

Stable URL: <https://www.jstor.org/stable/40231713>

Accessed: 27-08-2018 05:42 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Canadian Journal of Philosophy is collaborating with JSTOR to digitize, preserve and extend access to *Canadian Journal of Philosophy*

Hume on Responsibility and Punishment

PAUL RUSSELL
University of British Columbia
Vancouver, BC
Canada V6T 1W5

In this paper I pursue two closely related objectives. First, I articulate and describe the nature and character of Hume's theory of punishment. Second, in light of this account, I offer an assessment of the contemporary interest and value of Hume's theory. Throughout my discussion I emphasize the relevance and importance of Hume's views on moral responsibility to his account of punishment.¹ More specifically, I argue that Hume seeks to develop an account of punishment on the foundation of a naturalistic theory of responsibility – that is, a theory which draws our attention to the role of moral sentiment in this sphere.² Although this naturalistic aspect of Hume's theory of punishment has been largely overlooked by commentators, I maintain that it is, nevertheless, precisely this aspect of Hume's theory that is especially interesting from a contemporary perspective.

1 My concern with Hume's theory of responsibility in this context will be largely subservient to my primary concern which is Hume's general theory of punishment. I will not, accordingly, provide any detailed analysis of Hume's theory of responsibility. For a more detailed account of Hume's views on responsibility see my 'On the Naturalism of Hume's "Reconciling Project",' *Mind* 92 (1983), 593-600. Of related interest see also 'Causation, Compulsion and Compatibilism,' *American Philosophical Quarterly* 25 (1988), 313- 21.

2 There are, in general, two closely related elements involved in Hume's naturalism. The first is that which concerns Hume's project of 'science of man' (i.e. his empirical, scientific investigation into the principles of human nature). The second is that which concerns Hume's emphasis on the role of feeling in human life. Both these elements constitute major themes in Hume's philosophy and, as I will show, they both play an important role in shaping Hume's approach to the problem of punishment. On Hume's naturalism in general see, e.g., Barry Stroud, *Hume* (London: Routledge & Kegan Paul 1977), esp. Chs. 1 and 10. See also P. F. Strawson, *Skepticism and Naturalism: Some Varieties* (London: Methuen 1983), esp. Ch. 1, sect.3.

I

It would seem reasonable to assume that *if* Hume does have a clear position on the subject of punishment then it most likely is to be found in his major works the *Treatise* and the *Enquiries*.³ When we turn to these works, however, we find that what Hume has to say on this subject is both fragmented and cursory.⁴ His brief remarks on this issue largely emphasize the social utility of punishment. Punishment, he claims, is justified primarily on the ground that it motivates men to obey those rules of justice which are essential to 'the peace and security of human society.'⁵ Clearly, then, punishment must be justified, at least in part, in terms of a 'forward-looking' rationale. Given this, it would seem safe to conclude that Hume unambiguously rejects *pure* retributivism. That is, punishment cannot, on this account, be justified simply on the basis of guilt or wrongdoing, without any reference to consequentialist considerations.

This general interpretation of Hume's position on punishment certainly accords with the views of Hobbes and Schlick – two thinkers with whom Hume is closely associated in this context. More specifically, like Hume, Hobbes and Schlick are primarily concerned to emphasize the social utility of punishment. Both argue that punishment without any view to reform or deterrence is simply an act of 'hostility'

3 References will be to the Selby-Bigge edition of Hume's *Treatise of Human Nature*, revised by P.H. Nidditch (Oxford: Oxford University Press 1978); to the Selby-Bigge edition of the *Enquiries*, also revised by P.H. Nidditch (Oxford: Oxford University Press 1975); and also to *Essays Moral, Political and Literary*, T.H. Green and T.H. Grosse, eds., 2 vols. (London: Longmans 1875). The *Treatise* is abbreviated as THN; the first and second *Enquiries* are abbreviated as EHU and EPM respectively.

4 See, e.g., THN, 410: 'Tis indeed certain...'; EHU, 97-8: 'All laws being founded...'; THN, 609: 'Men have observ'd...'; EPM, 187: 'When any man....' It is an interesting question why Hume has relatively little to say on the subject of punishment. Suffice it to say, however, that for Hume, it would seem, it is the sense of honour and shame (i.e., pride or humility) that attaches to virtue and vice, rather than the prospect of reward or punishment, which is the principal motive to virtue. In other words, it is how we feel about ourselves, rather than our beliefs regarding how others will treat us, which is the sure path to virtue.

5 Cf. THN, II, 2, esp. sect. 7. Some commentators, probably with these passages in mind, interpret Hume's views on punishment as being fundamentally utilitarian in nature and thus 'as necessitating the rejection of desert' (John Kleinig, *Punishment and Desert* [The Hague: Nijhoff 1973], 49). By and large, however, commentators on Hume's philosophy have little or nothing to say on this subject. We find, in other words, that the thinness of Hume's remarks on this subject is reflected in the relevant secondary literature.

or 'vengeance.'⁶ It is important to note, however, that Hobbes's theory of punishment is not *purely* utilitarian in nature. According to Hobbes, punishment is, by definition, for a 'transgression of the law' and thus (necessarily) has a retributive element. Although Schlick's remarks are less clear on this point it would seem that he is willing to dispense with all retributive elements from punishment. To the extent that Hume's views are identified with those of Schlick, therefore, he would appear to be committed to a (purely) utilitarian theory of punishment.

There is, beyond these specific considerations, a more general explanation for why Hume is associated with Hobbes and Schlick on this subject. Hume is widely regarded as a major spokesman for the empiricist-compatibilist position in the free will dispute.⁷ As such, he is viewed as a key figure in a tradition that stretches from Hobbes to Schlick.⁸ Although Hume's remarks on the subject of punishment are rather slight, both Hobbes and Schlick do indicate, in some detail, the relevance of their views on free will to the problem of punishment.⁹ In light of the *supposed* parallels between Hume and Hobbes and Schlick

6 See, in particular, Hobbes's *Leviathan*, C.B. MacPherson, ed. (Harmondsworth: Penguin 1968 [1651]), Ch. 28; Schlick, 'When is a Man Responsible?', reprinted in B. Berofsky, ed., *Free Will and Determinism* (New York: Harper & Row 1966), 60.

7 That is to say, most philosophers would accept John Mackie's claim that Hume articulates the 'classic defence of compatibilism' (J.L. Mackie, *Ethics* [Harmondsworth: Penguin 1977], 245). Hume's discussion of the free will problem appears at THN, II,3,1-2 and EHU, Sect. 8. For a standard or orthodox interpretation on Hume's views on free will see, for example, Barry Stroud, *Hume*, ch.7.

8 Thus Isaiah Berlin, for example, speaks of the 'Hobbes-Hume-Schlick doctrine': *Four Essays on Liberty* (Oxford: Oxford University Press 1969), xv. Similarly, Jonathan Glover argues that 'almost identical versions of this doctrine [sc. compatibilism] are to be found in Hobbes, Hume, Mill, Russell, Schlick and Ayer': *Responsibility* (London Routledge & Kegan Paul 1970), 50n.

9 Hobbes, 'Of Liberty and Necessity,' selections reprinted in D.D. Raphael, ed., *British Moralists: 1650-1800* (Oxford: Oxford University Press 1969), 63-5; Schlick, 'When is a Man Responsible?', 59-61. The fundamental point which both Hobbes and Schlick are concerned to make in this context is that rewards and punishments serve to *cause* people to act in some ways rather than others and are, hence, of considerable social utility (cf. Hume, T,410). Schlick goes so far as to analyse responsibility in these terms. That is to say, on his account the issue of responsibility is to be interpreted in terms of the question regarding who it is of utility for us to punish (60). (A similar 'pragmatic' or 'forward-looking' approach is pursued in P. Nowell-Smith, 'Freewill and Moral Responsibility,' *Mind* 57 [1948], 45-61; and J.J.C. Smart, 'Free Will, Praise and Blame,' reprinted in G. Dworkin, ed., *Determinism, Free Will and Moral Responsibility* [Englewood Cliffs, NJ: Prentice Hall 1970], 208-13.)

on the subject of free will it seems reasonable, on the face of it, to interpret Hume's position on punishment as being, roughly, that of his like-minded colleagues in the empiricist-compatibilist tradition. In other words, from this general perspective we are encouraged to view Hume's account of punishment as being, essentially, utilitarian in nature.¹⁰

While there is, evidently, a strong utilitarian element in Hume's account of punishment, it seems equally clear that his remarks on this subject are not unequivocally utilitarian in nature. Moreover, it may be argued that there are several passages in the *Treatise* which lend support to an account of punishment that is essentially retributivist, rather than utilitarian, in nature. Hume suggests, for example, that in regard to *divine* laws, God may act not only in his 'magisterial capacity,' with a view to producing obedience, but also 'as the avenger of crimes *merely on account of their odiousness and deformity*' (T, 410-11, my emphasis). In several other passages Hume further suggests that we are naturally or psychologically disposed to retributive attitudes and practices. More specifically, he argues that our moral sentiments (qua calm forms of love and hate) naturally give rise to 'benevolence or anger; that is, with a desire of making happy the person we love, and miserable the person we hate' (T, 591).¹¹ In this way, we find that Hume's remarks in the *Treatise* and *Enquiries* lend some support to *both* (backward-looking) retributivist and (forward-looking) utilitarian interpretations. Nor is it clear how these divergent elements of his account of punishment are supposed to merge together into one unified theory. These tensions in Hume's position have led some commentators to conclude that 'it is not easy to say where Hume stood' on this issue.¹² A more severe critic might argue that Hume simply oscillates between utilitarian and retributivist views on punishment and he has, therefore, *no* clear position for us to articulate. In order to resolve this

10 I have argued elsewhere that this general perspective on Hume on free will is, in important respects, mistaken. (See the papers referred to in note 1.) In this paper I show that this mistaken perspective on Hume leads directly to confusion about his views on punishment.

11 See, more generally, THN, II, 2, 6. I follow Pall Ardal in thinking that Hume holds that the moral sentiments are calm forms of love and hatred. See his *Passion and Value in Hume's Treatise* (Edinburgh: Edinburgh University Press 1966), Ch. 6. Other relevant passages can be found at T, 348, 418 and 439; EPM, 302 and also *Essays*, Vol. 2, 139.

12 D.D. Raphael, 'Hume and Adam Smith on Justice and Utility,' *Proceedings of the Aristotelian Society* 73 (1973), 99; reprinted in *Justice and Liberty* (London: Athlone 1980).

problem we must, I suggest, look elsewhere in the corpus of Hume's writings.

It is, perhaps, surprising to find that Hume's most coherent remarks on the subject of punishment appear in his posthumously published essay 'Of the Immortality of the Soul.'¹³ In the second section of that essay Hume turns his attention to the 'moral arguments' which are supposed to prove the immortality of the soul. These arguments, he notes, are 'chiefly those derived from the justice of God, which is supposed to be further interested in the further punishment of the vicious and reward of the virtuous.' In this context Hume is especially concerned to argue that responsibility and punishment can be understood only within the fabric of human nature and human society. To this end Hume launches a number of arguments against the orthodox theological position. For our purposes the following points are the most important.

1. Hume maintains that the distribution of rewards and punishments is determined on the basis of our (human) moral sentiments. He writes:

By what rule are punishments and rewards distributed? What is the Divine standard of merit and demerit? Shall we suppose that human sentiments have place in the Deity? How bold that hypothesis!... To suppose measures of approbation and blame different from the human confounds everything. Whence do we learn that there is such a thing as moral distinctions, but from our own sentiments? (402)

In 1740, in a letter to Francis Hutcheson, Hume touches on the same point. He states: '... since Morality, according to your Opinion as well as mine, is determin'd merely by Sentiment, it regards only human Nature and human Life. This has often been urg'd against you, & the Consequences are very momentous. ...If Morality were determined by Reason, that is the same to all rational Beings: But nothing but experience can assure us, that Sentiments are the same. What Experience have we with regard to superior Beings? How can we ascribe to them any sentiments at all?'¹⁴

In this way, Hume is arguing that the moral sentiments serve as the foundation upon which we distribute rewards and punishments. Given that we have no knowledge of God's sentiments – or even if He has any sentiments – it is quite impossible for us to make any sense of

13 *Essays*, Vol. 2, 399-406

14 Hume repeatedly points out the difficulties involved in attributing human passions or sentiments to God: *Dialogues Concerning Natural Religion*, N.K. Smith, ed., 2nd ed. (Edinburgh: Nelson 1947), 156, 212, 226.

the basis upon which God is supposed to distribute rewards and punishments in a future state.

2. Hume argues that the point or purpose of punishment is that it secures or produces benefits for human society. He writes:

Punishment, without any proper end or purpose, is inconsistent with *our* ideas of goodness and justice; and no end can be served by it after the whole scene [sc. this world] is closed. ...What man, who has not met with personal provocation (or what *good-natur'd* man who has), could *inflict* on crimes, from the sense of blame alone, even the common, legal, frivolous punishments? And does anything steel the breast of judges and juries against the sentiments of humanity but reflections on necessity and public interest? (402; Hume's emphasis)

In other words, Hume maintains that a future state in which the wicked are punished would serve no end or purpose and thus would lack any proper or appropriate justification.¹⁵ The only proper end or purpose of punishment, he holds, is that it serves the interests of human society.

3. Hume points out that punishment must not be disproportionate to the offence:

Punishment, according to *our* conception, should bear some proportion to the offence. Why then eternal punishment for the temporary offences of so frail a creature as man? ...The chief source of moral ideas is the reflection on the interests of human society. Ought these interests, so short, so frivolous, to be guarded by punishments, eternal and infinite? The damnation of one man is an infinitely greater evil in the universe, than the subversion of a thousand millions of kingdoms. (402-3; Hume's emphasis)

Clearly, then, Hume is concerned to show that Divine punishment, being 'infinite and eternal,' is bound to be grossly disproportionate even when inflicted upon the worst of criminals.

In this way, Hume launches what is essentially a two-pronged attack on the theological position. First, given that we have no evidence

15 Of course, this is a separate issue from the question concerning the *utility* of (religious) *belief* in the existence of such a future state. That is to say, as Hume recognized, many would claim that we have utilitarian reasons to promote and encourage belief in the doctrine of future rewards and punishments. In reply to this (distinct) claim Hume maintains that beliefs of this general nature are, at best, of little social utility and that they in fact tend to undermine our natural human commitment to morality and the laws of society (D, 219-24; see also THN, 113-15; EHU, 147).

of the nature or existence of God's sentiments, there is no accounting for the basis upon which we can be responsible to Him and, therefore, there is no accounting for the basis upon which He distributes rewards and punishments. Second, in so far as we suppose God to distribute eternal punishments to those He disapproves of, such Divine punishment: (a) lacks any point or purpose since it produces no benefits for human society; and (b) given that it is infinite in duration, it will inevitably be excessive in the extreme. It would seem to follow, therefore, that according to Hume the theological doctrine of eternal rewards and punishments in a future life is from one point of view unintelligible, and, from another, morally repugnant.

Given this more elaborate statement of Hume's views on punishment how should we characterize the general nature of his position? There are, I believe, two features of Hume's discussion which merit particular attention. These are: (a) his 'mixed' or teleological retributivist account of punishment; and (b) the 'naturalistic' theory of responsibility which serves as the foundation of his teleological retributivism. I will begin by describing in more detail the nature of Hume's teleological retributivism.

Our analysis of Hume's remarks on punishment suggest that we must approach this issue by way of distinguishing several different questions, each of which requires, on Hume's account, a different answer.¹⁶ More specifically, there are at least two significantly different questions which Hume appears to be addressing himself to in this context. The first asks: What is the end or purpose of punishment? In other words, what is the general justifying aim of punishment considered as an institution or practice? The second asks: On what basis are punishments to be *distributed*? That is, who may be punished (i.e. who is *liable* to punishment) and how severely may we punish them (i.e. what *amount* of punishment are they liable to)? Hume, as I have noted, provides us with different answers to these different questions. According to his account, therefore, it is a mistake to assume that the justification of punishment involves an appeal to a single value or aim (e.g. deterrence, retribution, reform, etc.). In addressing ourselves to the issue of the general justifying aim, for example, we must appeal to considerations of social utility (i.e. deterrence and reform). Such (forward-looking) considerations, however, are not relevant or appropriate when we turn to the question of the distribution of punishment

16 My approach here draws heavily on H.L.A. Hart's influential statement of the teleological retributivist (i.e. 'mixed' or 'reconciliationist') theory of punishment in 'Prolegomenon to the Principles of Punishment,' reprinted in *Punishment and Responsibility* (Oxford: Oxford University Press 1968), 1-27.

(i.e. liability and amount). Here we must take note of (backward-looking) retributivist considerations. Considerations of this nature, Hume suggests, must be interpreted in terms of our moral sentiments.

The moral sentiments, it is claimed, provide us with a framework within which it may be determined who *deserves* to be punished and who does not. That is, according to this account, there is an intimate connection between the question of whether or not we deem someone liable to punishment and the question of whether or not that person arouses the appropriate negative moral sentiments in us. Someone who does not arouse such sentiments in us when we contemplate his actions or character cannot be justly punished *even if* the infliction of punishment would, all things considered, produce beneficial consequences. In other words, it is a necessary condition of being justly punished that you are also justly disapproved of or blamed. Further, the *degree* to which a person whom we disapprove of is liable to punishment will clearly be limited by the degree to which we disapprove or blame him (i.e. the amount of punishment must not be out of proportion to the strength of our disapproval).¹⁷ In this way, we find that our moral sentiments account for the principles which govern and limit the distribution of punishment.

Hume maintains that punishment is justified only if it serves the public interest. It would seem to follow, therefore, that where punishment does not serve the public interest it cannot be justified *even if it is deserved*. That is to say, from the fact that someone arouses negative moral sentiments in us it does not straightforwardly follow that we are justified in punishing him. It follows only that he is *liable* to the *appropriate degree* of punishment. On this account, then, the question of whether or not the man *deserves* to be punished should be kept distinct from the question of whether or not the infliction of punishment would produce some beneficial consequences.¹⁸ It is, accordingly, especially important to note that on this view the general justifying aim of punishment cannot be accounted for simply in terms of our negative moral sentiments towards the individual(s) concerned. While our moral sentiments may serve as an adequate basis upon which to account for the distribution of punishment, they do not, of themselves,

17 Cf. THN, 348: 'Nothing is more evident'

18 Hume's position on this issue may be contrasted with Austin Duncan-Jones's account of Butler's position. He reports Butler as taking the view that the notion of desert suggests '*both* that a man's conduct is bad *and* that to penalize him is useful, or *both* that his conduct is good *and* to praise or reward him is useful' (*Butler's Moral Philosophy* [Harmondsworth: Penguin 1952, 140; my emphasis]).

provide us with any adequate account of the *purpose* or *point* of punishment. It would be a mistake, in other words, to attempt to justify the practice of punishment by referring solely to our (negative) moral sentiments without making any reference to the utility of such a practice. To this extent Hume's theory is not committed to, and does not license, any crude form of retributivism. Should we find, for example, that the institution or practice of punishment lacks any satisfactory teleological rationale (i.e. it is of no social utility) then Hume may consistently maintain that such an institution or practice should be suspended or abandoned altogether.

Hume's position on this issue may be further clarified by way of distinguishing between two very different modes of retributivism. That is to say, it is crucial that we distinguish between 'negative retributivism,' the doctrine that we must not punish the innocent (nor excessively punish the guilty), and 'positive retributivism,' the doctrine that we have an obligation to punish the guilty. Our moral sentiments towards an individual do not, of themselves, generate an obligation to punish that individual. That is to say, as I have noted, it does not follow from the fact that an individual *deserves* to be punished that we are, necessarily, justified in punishing him (as such punishment may be of no social utility). Hume's theory, therefore, does not lend any support to 'positive retributivism.' It does, however, account for the 'negative' retributive principle which is essential to the teleological retributivist position. That is, our moral sentiments, by way of determining the distribution of punishment, ensure that the innocent must not be punished, nor the guilty excessively punished. Similarly, on the other side of the same coin, our moral sentiments also enable us to determine who we are *permitted* to punish (i.e. who is liable to punishment). In this way, the moral sentiments provide what may be described as the 'negative' rationale of punishment. They indicate the *limits* within which the attitudes and practices associated with punishment must take place. Theories and systems of punishment which try to ignore the limits imposed by moral sentiments are, on this account, fundamentally flawed. The framework which the moral sentiments impose upon any system of punishment cannot be ignored because this framework constitutes the very foundation upon which any adequate or plausible theory of punishment must be erected.

It is widely held that Hume's theory of justice is a form of rule-utilitarianism. It should be clear, however, that as regards punishment Hume's position is more complex than this. In particular, as I have emphasized, the principles governing the distribution of punishment reflect considerations of desert, interpreted in terms of moral sentiment, rather than those of utility. In this regard, the obvious point of comparison is John Rawls's well-known rule-utilitarian approach to

punishment.¹⁹ According to this account, no rule or practice permitting punishment of the innocent (i.e. 'telishment') is likely to have any utilitarian justification. In this way, for the rule utilitarian it is considerations of social utility which are the ground of our prohibition against punishing the innocent. We misrepresent Hume's analysis of the retributive element of punishment if we interpret him along these lines.

Moral sentiment, evidently, plays a central role in Hume's account of punishment. This is indicative of the more general point that Hume is committed to a naturalistic account of moral responsibility.²⁰ One of the objectives of Hume's 'science of man' was to *describe* the circumstances in which men are *felt* to be responsible. This description is given largely in terms of Hume's associationist psychology and the mechanism of the indirect passions – the moral sentiments being understood as calm forms of love and hatred. The details of this account do not concern us in this context. The crucial point is that, for Hume, responsibility must be understood in terms of our natural psychological reactions to the moral qualities and character traits of our fellow human beings. We find that the actions and character traits of other people inevitably generate emotional responses and reactions in us (i.e. moral sentiments).²¹ Clearly, then, on this account, responsibility must be viewed as part-and-parcel of our emotional make-up and, as such, a 'given' of our human nature.²² According to this view, therefore, a proper understanding of the nature and circumstances of responsibility requires that we develop a clearer understanding of the role that moral sentiment plays in human life.²³

19 John Rawls, 'Two Concepts of Rules,' reprinted in P. Foot, ed., *Theories of Ethics* (Oxford: Oxford University Press 1967), 144-70

20 A naturalistic interpretation of Hume's views on responsibility is defended in my 'On the Naturalism of Hume's "Reconciling Project".' A more detailed account of Hume's description of the 'mechanism of responsibility' is provided in that paper.

21 In this way, while we may, on Hume's account, suspend or abandon the institution or practice of *punishment*, there is no question of us abandoning or suspending the whole framework of the moral sentiments.

22 See T,474: 'These sentiments are so rooted in our constitution and temper...'; EPM,273: 'But the sentiments of humanity...'; EPM,322: 'Every one may employ *terms* in what sense he pleases....'

23 Hume's views regarding the relevance of freedom to ascriptions of responsibility must be understood within this framework. When an agent acts according to the determination of his will (i.e. when he enjoys 'liberty of spontaneity') it is possible for us to infer his character from his action. In these circumstances our moral sentiments are naturally aroused. By contrast, where and when an agent is

It is this naturalistic account of responsibility which, according to Hume, must serve as the foundation of any adequate theory of punishment. In particular, it is essential, on this account, that we interpret the retributive features of punishment in terms of our moral sentiments. This important feature of Hume's discussion of punishment (i.e. his concern with the moral sentiments) is entirely absent from Hobbes's and Schlick's discussions. To this extent, therefore, it seems evident that Hume develops his theory along fundamentally different lines from those pursued by Hobbes and Schlick.²⁴ Moreover, whatever objections Hume's theory is vulnerable to, it is *not* vulnerable to the standard objection levelled against purely utilitarian theories: namely, that they license the punishment of the innocent or the excessive punishment of the guilty. Suffice it to note that Schlick's account is widely thought to be vulnerable on precisely this point.²⁵

II

Hume has not been alone in his efforts to construct a theory of punishment on the basis of naturalistic foundations. On the contrary, Hume's friend and fellow Scot, Adam Smith, developed a theory of punishment along very similar lines. Indeed, on the face of it, Smith's position is very close to that of Hume's, in so far as he too is concerned to *describe* the moral psychology on which our practices of reward and punishment rest.²⁶ In light of this, it may be suggested that Hume's

subject to external force or violence (i.e. lacks liberty of spontaneity) or his actions are due to 'indifference' (i.e. are uncaused) it is impossible for us to infer his character from his action. It is, according to Hume, a matter of psychological fact that in these circumstances no moral sentiment would be aroused in us and, hence, the agent would not be regarded as responsible.

24 In light of these observations it may be argued that in so far as Hume's views on *punishment* have been misunderstood and misrepresented the root difficulty lies with confusion about his views on *responsibility*. More specifically, the general tendency to place Hume squarely in the empiricist-compatibilist tradition, alongside Hobbes and Schlick, has led commentators to overlook Hume's concern with the role of moral sentiment in this sphere. Clearly, as our analysis reveals, without an adequate interpretation of Hume's views on responsibility it is impossible to develop a proper understanding of his position on punishment.

25 See, e.g., C.A. Campbell's objections to 'Schlick's essentially "forward looking" interpretation of punishment and responsibility': 'Is "Freewill" a Pseudo-Problem?', reprinted in Berofsky, *Free Will and Determinism*, 112-18.

26 Smith, *The Theory of Moral Sentiments*, D.D. Raphael and A.L. Macfie, eds. (Oxford: Oxford University Press 1976); abbreviated as TMS. Also relevant is Smith, *Lectures on Jurisprudence*, R.L. Meek, D.D. Raphael, and P.G. Stern, eds. (Oxford:

theory of punishment is best understood as being simply an (under-developed) variant of the sort of (naturalistic) theory that Smith describes. Such an interpretation, however, would be mistaken. The differences between Hume's and Smith's theories are at least as instructive as their similarities.

According to Smith, we regard punishment as just and deserved in those circumstances where an impartial spectator would sympathize with (i.e. share) the resentment felt by the injured party.²⁷ More precisely, an action deserves punishment when 'it appears to be the proper object of resentment' (TMS, 67). It is the sentiment of resentment 'which most *immediately and directly* prompts us to punish.' To punish, Smith maintains, is simply 'to return evil for evil that has been done' (TMS, 68). In this way, we find that an offender is deemed 'the proper object of punishment, when we thus entirely sympathize with, and thereby approve of, that sentiment which prompts us to punish' (TMS, 74). In other words, Smith claims that in 'going along' with the sentiment of resentment from which the action of punishment 'proceeds,' we must necessarily *approve of the punishment*, and regard the person against whom it is directed as its proper and suitable object.

Smith's retributivism is in no way half-hearted. An action surely deserves punishment, he says, 'which every body who hears of it is angry with, and upon that account *rejoices* to see punished' (TMS, 70; my emphasis). Smith continues in this vein:

When we see one man oppressed or injured by another, the sympathy which we feel with the distress of the sufferer seems to serve only to animate our fellow-feeling with his resentment against the offender. We are rejoiced to see him attack his adversary in his turn, and eager and ready to assist him whenever he exerts himself for defence, or even for vengeance within a certain degree. (TMS, 70-1)

These retributivist sentiments are given full vent by Smith when he turns to consider our reaction to murder and our sympathy with the victim.

Oxford University Press 1978); abbreviated as LJ. Smith states (TMS, 77): 'Let it be considered too, that the present inquiry is not concerning a matter of right, if I may say so, but concerning a matter of fact. We are not at present examining upon what principles a perfect being would approve of punishment of bad actions; but upon what principles so weak and imperfect a creature as man actually and in fact approves of it.'

²⁷ In general, on Smith's account we sympathize with another person when we imaginatively place ourselves in his position and feel what he feels (i.e. we find that our sentiments 'correspond' with those of the person involved). In these circumstances we regard his sentiments as proper or appropriate to their object, and thus we approve of them.

We feel that resentment which we imagine he ought to feel.... His blood we think *calls aloud for vengeance*. The very ashes of the dead seem to be disturbed at the thought that his injuries are to pass unrevenge.... And with regard, at least, to this most dreadful of all crimes, *Nature, antecedent to all reflections upon the utility of punishment, has in this manner stamped upon the human heart, in the strongest and most indelible characters, an immediate and instinctive approbation of the sacred and necessary law of retaliation.* (TMS, 71; my emphasis)

In light of passages such as this, it would be hard to overstate the strength of Smith's (positive) retributivism. The contrast with Hume, in this respect, is perfectly clear.

While Smith, unlike Hume, rejects any role for forward-looking consequentialist considerations in *justifying* punishment he, nevertheless, does not entirely overlook such considerations.²⁸ A number of philosophers, he notes, have suggested that it is consideration of the public good which is 'the real source of the punishment of crimes.'²⁹ Smith, as we have seen, rejects this suggestion. However, he does acknowledge that there is an *indirect* connection between our retributive practices and the welfare of society. Indeed, Smith points out that there can be no doubt 'that we frequently have occasion to confirm *our natural sense of the propriety and fitness of punishment*, by reflecting how necessary it is for preserving the order of society' (TMS, 88; cf. TMS, 77: 'The very existence of society...'). Nevertheless, be this as it may, 'The Author of nature,' Smith suggests, 'has not entrusted it' to man's reason to find out that a certain application of punishments is the proper means of attaining the welfare and preservation of society. In order to secure the preservation of society – an end which all men naturally desire – nature has not only 'endowed mankind with an appetite for the end which she proposes, but *likewise with an appetite for the means by which alone this end can be brought about, for their own sakes, and independent of their tendency to produce it*' (TMS, 77; my emphasis). In other words, according to Smith we naturally and instinctively approve of those punishments which secure an end which we also, independently,

28 Smith's views on this issue, it should be noted, are not entirely uniform. More specifically, Smith does say that on some occasions 'we both punish and approve of punishment, *merely from a view to the general interest of society*, which, we imagine, cannot be otherwise secured' (TMS, 90; my emphasis, and also LJ, 105). It is not entirely obvious, however, how we are supposed to interpret such exceptional cases; nor how Smith's remarks in this direction square with his analysis of the standard case. On this see Raphael, 'Hume and Adam Smith', 96-7.

29 LJ, 104. Here Smith probably has Hume primarily in mind – although it is 'Grotius and other writers' to whom he actually refers. See the editors' remarks at TMS, 87-8.

desire: viz. the welfare of society. On this basis, Smith argues that 'revenge of the injured' will 'regulate' our practice of punishment in such a way that it (indirectly) secures precisely those ends (i.e. deterrence, reform, etc.) which philosophers such as Hume have *mistakenly* suggested are the source of our 'original interest' in the punishment of crimes (LJ, 105; TMS, 89).³⁰

It seems clear that despite their obvious similarities there are, nevertheless, significant differences between Hume and Smith on the subject of punishment. There are three particularly important general points to be noted.

1. Both thinkers are concerned to develop a theory of punishment on the basis of a naturalistic account of the moral sentiments. More specifically, for both Hume and Smith it is crucial that we interpret the retributive features of punishment in terms of our moral sentiments. In this respect their approach to the problem of punishment is fundamentally different from that of Hobbes and Schlick.

2. Hume, following Hobbes, and unlike Smith, develops a 'mixed' or teleological retributivist theory of punishment. Both Hobbes and Hume are agreed that any adequate theory of punishment must appeal to utilitarian considerations as the general justifying aim of punishment. Neither, accordingly, would accept Smith's form of positive retributivism.

3. Hume, along with both Hobbes and Smith, rejects any *purely* utilitarian approach to punishment. For all three thinkers there is an ineliminable retributive element involved in punishment.

It seems clear, in light of these observations, that Hume's account of punishment differs in significant respects from both that of Hobbes and Smith. We may conclude, therefore, that Hume has a theory of punishment which is not only distinctive, but also far richer and more complex than has hitherto been recognized.³¹

30 It may be noted that Smith's claims regarding the *indirect* utility of retributive practices are highly questionable. That is, Smith claims that the 'law of retaliation' will effectively 'regulate' punishment in such a way that it indirectly secures our utilitarian ends (deterrence, reform, etc.). At least two difficulties arise for this claim: (a) in many cases such practices may serve no such ends – they may even be counter-productive; (b) from a utilitarian perspective there may be *more effective* alternative strategies available to us.

31 Within the confines of this paper it is not possible to explore the significance which this general interpretation of Hume's theory has for the more specific or narrower remarks that Hume makes on the subject of punishment. Two passages,

III

In my introductory remarks I suggested that the most interesting aspect of Hume's discussion, from a contemporary perspective, lies in his effort to develop a theory of punishment on naturalistic foundations (i.e. his effort to account for the role that moral sentiment plays in this sphere). The contemporary significance of this aspect of Hume's discussion can be better appreciated if we compare and contrast Hume's and Smith's views in light of more recent literature on this subject.

Behind the specific issues separating Hume and Smith lies the deeper and wider problem of how we are to understand the relationship between responsibility and punishment. Both Hume and Smith, as I have indicated, are agreed that punishment must be understood as resting upon, or developing out of, the fabric of our moral sentiments. Beyond this, however, there is deep disagreement about how we should interpret the implications of a naturalistic theory of responsibility for the problem of punishment. This is an issue which has recently resurfaced in a particularly prominent context: namely, in the context of Strawson's highly influential statement of the naturalistic approach to responsibility.³² The similarities between Hume's and Strawson's

however, merit brief comment: (1) At T,608-9 Hume, notoriously, suggests that natural abilities and moral qualities are 'on the same footing' as regards their tendency to arouse moral sentiments in us. In other words, according to Hume people are held accountable or responsible for both. There is, nevertheless, he suggests, a distinction to be drawn between them in so far as our moral qualities, but not our natural abilities, 'may be chang'd by the motives of rewards and punishments, praise and blame.' Clearly there is much in this account to question. However, Hume's position on this more specific issue is, I suggest, more intelligible when considered within the framework of his general theory of punishment. (2) At EPM,322 – developing points raised from the passage at T,608-9 (taking issue, in particular, with the suggestion that the distinction between voluntary and involuntary be made the foundation of a theory of morals) – Hume objects to those philosophers and/or divines who treat 'all morals as on a like footing with civil laws, guarded by the sanctions or reward and punishment.' The general point that Hume is concerned with here is that it is crucial that we distinguish between situations where our moral sentiments have been legitimately aroused and circumstances in which we are justified in distributing rewards and punishments on the basis of these sentiments. This fundamental point is firmly embedded in Hume's general theory of punishment. Indeed, it lies at the heart of his objections to the theological doctrine of future rewards and punishments.

32 'Freedom and Resentment,' reprinted in G. Watson, ed., *Free Will* (Oxford: Oxford University Press 1982), 59-80. See also Strawson, *Skepticism and Naturalism: Some Varieties*, Ch.2. Unlike Hume (and Smith), Strawson speaks of 'reactive attitudes and feelings.' He notes, however, that it is 'a pity that talk of the moral sentiments has fallen out of favour' as that phrase would be a 'good name' for the network of attitudes and feelings he is concerned with (79). It should also

naturalistic approach to responsibility are in many respects quite striking. In general, the salient features of Hume's views on responsibility, as described above, reappear in Strawson's account. Strawson's remarks regarding the implications of his views for the problem of punishment are, nevertheless, both brief and obscure.³³ More specifically, as it stands, it is unclear whether his remarks should be interpreted as lending support to a form of positive retributivism, along the lines of Smith, or, alternatively, to a form of teleological retributivism along the lines of Hume. If it is the former theory which Strawson is embracing then it may be argued, for reasons which I will explore, that such an account is not satisfactory. More generally, it may be argued that until it is shown that a naturalistic account of responsibility can be rendered consistent with a plausible theory of punishment then such an account remains suspect.³⁴ In this way, it seems clear that the issues which separate Hume and Smith reappear at the very heart of Strawson's discussion. Granted the contemporary importance of Strawson's views on this subject we cannot, I suggest, afford to overlook Hume's and Smith's rival positions on this matter.³⁵

be noted in this context that Strawson's paper is in large measure a critique or attack upon Schlickean, 'forward-looking' accounts of responsibility and punishment.

- 33 'Savage or civilized,' Strawson claims, 'we have some belief in the utility of the practices of condemnation and punishment ...' (78). For Strawson, however, this is not the central issue. Rather, the point which Strawson is concerned to establish is that 'to speak in terms of social utility *alone* is to leave out something vital in our conception of these practices' (viz. our moral sentiments).
- 34 Even though Strawson's paper has generated a great deal of comment and criticism very little has been said about the problem of punishment as it arises for him. Instead, commentators have focused their attention, almost exclusively, on the narrower issues of freedom and responsibility. This situation is particularly surprising in light of the fact that Smith (in whose footsteps Strawson clearly follows) devotes a great deal of his attention to the problem of punishment. It is fair to say, I believe, that this is indicative of the extent to which Strawson's naturalistic theory of responsibility has been cut off from its historical roots.
- 35 While Hume's views on the subject of punishment have received little attention over the years Smith's naturalistic account of positive retributivism has, by contrast, succeeded in attracting a number of followers. The most important of these have been Edward Westermarck, *Ethical Relativity* (London: Kegan Paul 1932; reprinted Westport, CT: Greenwood 1970); see esp. Ch. 3, and, more recently, John Mackie, 'Morality and the Retributive Emotions,' reprinted in *Persons and Values* (Oxford: Oxford University Press 1985), 206-19. The Smithian flavour of Mackie's general position is quite obvious – indeed, it can be directly traced to Smith via Westermarck. In order to preserve the coherence and direction of my own discussion, however, I will continue to focus attention on Smith's views.

My analysis of the contrast between Hume's and Smith's views will focus on two closely related issues: (i) the relationship between moral sentiments and the justification of punishment; and (ii) the relationship between retributive practices and the will.

(i) Hume, along with Smith, takes the view that it is impossible to develop a convincing or plausible 'forward-looking' or consequentialist rationale for our moral sentiments. That is, both are agreed that our moral sentiments are essentially retrospective in nature. Moreover, both are agreed that such sentiments arise naturally and spontaneously in reaction to the perceived moral qualities of another person.³⁶ However, Hume also maintains, contrary to Smith's view, that a wholly 'backward-looking' or retributive account of *punishment* would be equally inadequate and implausible. The rationale of punishment, it is argued, raises separate issues from those which arise solely within the framework of our moral sentiments. That is to say, there exists what may be described as a 'justificatory gap' between our moral sentiments and our practice of punishment. The only way to bridge this gap is by taking note of the relevance of forward-looking, consequentialist considerations.

Where, according to Hume, does this 'justificatory gap' arise? Fundamental to Smith's position, as we have noted, is the claim that in blaming an individual (i.e. regarding him as an appropriate object of resentment) we inevitably or necessarily approve of the infliction of punishment upon that individual. In other words, according to Smith, if our sentiments of blame or resentment are appropriate and suitable to their object, then so too are those actions of punishment which are motivated by these sentiments. Punishment, it appears, requires no

36 Without going too deeply into the complexities of Hume's system two further points should be noted: (1) Some of Hume's remarks may be interpreted as suggesting that our moral sentiments are incapable of justification (e.g. T, 413-18, 455-70). Nevertheless, Hume makes it clear that our moral sentiments, like other passions, may be said to be *indirectly* reasonable or unreasonable, justified or unjustified, in so far as they arise out of true or false beliefs. (On the role of reason in morals see esp. EPM, Sect. 1.) (b) The beliefs which generate or give rise to the moral sentiments concern the supposed pleasurable or painful qualities of mind or character of the individual in question. Hume suggests that it is our expectations about the *tendencies* of an individual's character which shape or condition our moral sentiments. The grounds of such expectations are inevitably rooted in our past experience and knowledge of the action and behaviour of the individual. The crucial point, however, is that *our* moral sentiments are not to be justified in terms of *their* future consequences or utility. Whether or not such sentiments are of utility they will naturally arise in the relevant circumstances (see, e.g., EPM, 273 and 322). In this sense they are neither capable of, nor require, a consequentialist rationale.

further or independent justification than that which is already presupposed in viewing someone as a reasonable or appropriate object of blame or resentment. It is, therefore, a mistake to seek to provide some further or independent justification for inflicting punishment beyond reference to the appropriate or suitable nature of those sentiments which motivate such actions. Resentment and retribution, on this account, come in one indivisible package. To regard resentment as justified and appropriate is, ipso facto, to regard punishment as justified and appropriate.

It is precisely at this stage that Hume rejects Smith's model. It simply does not follow from the fact that our sentiments of blame or resentment are justified that we are therefore justified in intentionally inflicting suffering or pain on the individual(s) concerned.³⁷ The backward-looking, retributive considerations which serve to justify our moral sentiments cannot, of themselves, bridge this gap. At the very least, therefore, we require some further argument here to show how, or why, such considerations can be 'extended' to cover the further issue of punishment. Simply to make reference to the 'necessary law of retaliation' (TMS, 71, 82), or to the principle of 'returning evil for evil done' (TMS, 68), is either to provide no further justification at all, or, worse still, to misrepresent such practices as psychologically or practically speaking inescapable features of human life.³⁸

This 'gap' in Smith's account of punishment is most apparent when we focus attention on the institution or practice of punishment taken as a whole. That is to say, the obvious question to raise at this juncture is: Why should society spend its energies and resources on constructing and maintaining institutions and practices of this nature? Smith's response here, such as it is, seems sorely inadequate. In

37 Regarding the relationship between blame and punishment see, e.g., J.E.R. Squires, 'Blame,' reprinted in H.B. Acton, ed., *The Philosophy of Punishment* (London: MacMillan 1969), 204-11; and Richard Wasserstrom, 'Some Problems in the Definition and Justification of Punishment,' in A.I. Goldman and J. Kim, eds., *Values and Morals* (Dordrecht & Boston: Reidel 1978), 307-8. (Wasserstrom, I note in passing, raises several interesting objections against the sort of 'mixed' or teleological retributivist theory of punishment which I attribute to Hume. Suffice it to say, in this context, that I am not persuaded that his objections tell against Hume's position.)

38 Consider the following tale told about Zeno of Citium, a necessitarian, by Diogenes Laertius: 'We are told that [Zeno] was once chastising a slave for stealing, and when the latter pleaded that it was his fate to steal, "Yes, and to be beaten too," said Zeno' (Diogenes Laertius, *Lives of Eminent Philosophers*, 2 vols. [London: 1925], Vol.2, 135). In many respects, Smith's general position seems to be akin to that of Zeno.

respect of this issue it is wholly unconvincing to be told that the sole point or purpose of punishment is simply that it implements the 'necessary law of retaliation' and ensures that we return 'evil for evil done.' So described, punishment has no point or value at all.³⁹ Smith's efforts to account for the *indirect* utility of such practices, it may be argued, indicate that he is himself aware of the 'gap' in his account. This appeal to considerations of indirect utility is, however, beside the point. As I have already noted, it is according to Smith *irrelevant*, from a justificatory point of view, that our retributive practices turn out to be of (indirect) utility. Clearly, therefore, Smith and his followers remain entirely committed to the view that such institutions and practices are just and proper *irrespective of their consequences*.⁴⁰ Seen from the perspective of Hume's alternative views on this subject, however, this claim, so presented, is highly suspect. It is suspect because, as I have suggested, it leaves the 'justificatory gap' entirely unbridged.

Hume's complex theory of punishment makes it possible for us to see our way past an apparent paradox in this sphere. On the face of it there appears to be a conflict or tension between the fundamental tenets of naturalistic theory of responsibility and the claim that punishment requires a teleological element in its justification. More specifically, there appears to be a conflict between: (a) the claim – as expressed by Strawson – that 'the whole framework' of our moral sentiment requires no 'external rational justification' (as it is a 'given' of our human nature regarding which we have no choice); and (b) the claim – as expressed by Hart – that punishment requires some adequate 'general justifying aim' (as it is not a 'given' of human nature

39 Smith may be understood to be suggesting – in a more utilitarian vein – that the point of punishment is that it gratifies our desire for retribution. This line is pursued by James Fitzjames Stephen, *Liberty, Equality, Fraternity* (Cambridge: Cambridge University Press 1967 [1873]), 152. Stephen argues that common crimes are punished, 'not only because they are dangerous to society, but also for the sake of gratifying the feeling of hatred – call it revenge, resentment or what you will' It follows, Stephen suggests, that the criminal law is 'an emphatic assertion of the principle that the feeling of hatred and the desire for vengeance ... are important elements of human nature which ought in such cases to be satisfied in a regular public and legal manner.'

40 Consider, for example, the theological implications of Smith's account, and how it contrasts with Hume's position. That is, granted the ontology presupposed in the doctrine of a future state, Smith *cannot* follow Hume in regarding the existence of Hell as a barbarism without point or purpose. Indeed, Smith takes quite the opposite view. See TMS, 91 and 163f. (Although Smith would grant, no doubt, that Hell involves the *excessive* punishment of the guilty.)

and is, thus, a matter over which we *do* have choice).⁴¹ Smith and his followers would eliminate this apparent tension by rejecting the suggestion that punishment requires some general justifying aim. From their perspective our retributive practices, no less than the whole framework of our moral sentiments, are a given of human nature and, as such, are neither capable of nor require any 'external rational justification.'⁴² Hume's complex theory succeeds, I believe, in providing us with a more satisfactory solution to this problem. By drawing a sharper distinction between our moral sentiments and our retributive practices, and by offering an alternative account of the relations between them, the complex theory enables us to explain how it is that consequentialist considerations are called for with regard to the latter but not the former. That is, while it is essential, on this account, that *punishment* be justified in terms of consequences, no such rationale is required for the framework of our moral sentiments. Clearly, once these points are established, it is possible to reconcile the *prima facie* conflicting demands of a naturalistic account of responsibility with a teleological element in the justification of punishment. In general, an understanding of the significance of this gap between our moral sentiments and our retributive practices is crucial if we are to avoid drawing the mistaken conclusion that a naturalistic theory of responsibility necessarily commits us to some strong form of positive retributivism.

(ii) Our analysis of the 'justificatory gap' between blame (resentment) and punishment reveals a further point of disagreement between Hume and Smith. Lying behind their divergent views on the justificatory issue is a more fundamental disagreement concerning moral psychology. It is disagreement at this level which explains their more obvious disagreement at the level of justification. Smith's position involves two distinct, but intertwined, claims. First, as has been noted, Smith claims that punishment may be justified without reference to its consequences. Second, Smith also suggests – although this claim is not so clearly developed – that punishment is embedded in our human nature in such a way that we are left, in some important sense, with no choice on the matter. That is to say, Smith speaks of punishment as the 'natural consequence of resentment' (TMS, 79) and thus as something which is (for us), in the relevant circumstances, a practical or psychological necessity. In other words, for Smith retributive practices are,

41 Strawson, 'Freedom and Resentment', 78-9; Hart, *Punishment and Responsibility*, 8-11

42 This theme is pursued in some detail in Mackie, 'Morality and the Retributive Emotions.'

in some sense, 'built into' our human nature. 'As every man doth, so shall it be done to him, and *retaliation seems to be the great law which is dictated to us by nature*' (TMS, 83; my emphasis). This claim allows for two quite different interpretations. On the strong interpretation Smith may be read as arguing that in respect of our will our retributive practices are on the same footing as our (reactive) moral sentiments. That is, Smith may be understood to be claiming that our retributive practices are essentially involuntary or spontaneous in the sense that they are not a matter of decision or choice *at all*. A few of Smith's remarks lend themselves to this interpretation, perhaps, but on the whole it is clear that this (self-evidently implausible) view is not what he has in mind.⁴³

On the weak interpretation Smith may be taken to acknowledge or recognize that our retributive practices are indeed voluntary and, in some measure, depend on our choices or willings. That is, our retributive practices, on this account, are a matter of deliberation and choice in a way that our moral sentiments clearly are not. However, the scope for deliberation is, nevertheless, severely limited. In this respect Smith draws a useful analogy between our desire for retaliation and other 'original and immediate instincts' such as our desire for food, water and sex (TMS, 77-8). The general point of this analogy seems to be that our natural desire for retaliation is so fundamental to our human nature that it inevitably shapes our will and leads to action. It is inconceivable, in other words, given our human nature, that these desires could be rendered 'inert' in actual practice. While there may be some scope for choice in *shaping* and *directing* these practices there is, on this account, no question of us suspending or abandoning these practices altogether. Such practices, it is suggested, are simply too deeply embedded in our human nature for such an option to arise. Clearly this view of things blocks all proposals to eliminate – or even radically reform – our practice of punishment (i.e. on the ground that given our human nature such proposals are impracticable). In short, Smith is understood, on this account, to be arguing that we can no more choose to suspend or abandon our retributive practices as a whole than we can choose to abandon the whole framework of our moral sentiments. On the contrary, our natural human commitment to the latter ensures that we

43 Implausible as this strong view may be, several of Mackie's comments ('Morality and the Retributive Emotions,' 216-19) lend themselves to this interpretation. In general, Mackie blurs the distinction between retaliating without a view to consequences and retaliating spontaneously or instinctively without making any 'conscious choice' on the matter.

are inescapably committed to the former. Hence, within this sphere there are natural limits to what we may decide or choose to do.⁴⁴

The point at which Hume and Smith differ regarding the relation between the will and our retributive practices should now be quite clear. From a Humean perspective Smith's model of the way in which the practice of punishment is embedded in human nature substantially misrepresents the facts. The crucial point here is not simply that practices of this nature involve the intentional infliction of suffering and pain and hence must be *chosen* – this, as I have indicated, is a point which Smith can concede (i.e. on the weaker interpretation). Rather, the crucial point, from a Humean perspective, is that our commitment to the whole framework of the moral sentiments is not such that it is psychologically or practically impossible for us to free ourselves of retributive practices. That is to say, however natural retributive feelings and desires may be, we have no reason to regard our will as dominated and controlled by them. On the contrary, it is the sign of a civilized and humane mind (or society) that such retributive impulses are controlled and curbed in such a way that we can ensure that they serve only socially desirable ends.⁴⁵ Where they do not serve or secure such ends they must be checked, and to that extent suspended or abandoned where necessary. This difference in our practical situation in respect of punishment accounts for the fact that it requires a further *level* of justification (and that it is at this level that consequentialist or forward-looking considerations come into play). Above all, from a *practical* point of view we may not regard ourselves as 'victims' of our own moral sentiments. While we do not choose our moral sentiments, we do choose whom, and when, to punish. In this respect we stand in an utterly different position in relation to the practice of punishment than we do to our essentially spontaneous or involuntary emotional reactions to the moral qualities of other people. There is, then, on this account, sufficient scope for us to choose to suspend or abandon our retributive practices in circumstances where we can see that they secure no socially desirable ends.

44 So described, Smith's position may be (loosely) interpreted in terms of Harry Frankfurt's account of free will ('Freedom of the Will and the Concept of a Person,' in Watson, ed., *Free Will*). That is, our desire for retaliation is such that whatever our reflective attitude may be with respect to these desires they will nevertheless be effective in leading to action. In other words, it is impossible for us to restructure our will in such a way that our resentment does not lead to retaliation or retribution. We cannot alter or change our will in this respect. This is consistent with the (obvious) fact that retaliatory actions are done of our own will and thus we enjoy freedom of action in respect of them.

45 Clearly, for Hume, someone who is willing to intentionally inflict pain and suffering for no purpose other than vengeance is simply vicious.

I believe that there is much to be said for this Humean reply to Smith. Nevertheless, I have one or two reservations which I will briefly describe. First, even if we concede that it is possible for us to 'check' our desire for retaliation in light of utilitarian considerations, such a policy seems to encourage and condone widespread hypocrisy and insincerity. That is to say, if we are constantly deciding whether or not to *express* our moral sentiments, in word or deed, on the basis of (forward-looking) utilitarian considerations then it seems clear that we will frequently have reason *not* to express our true or sincere sentiments towards one another. In these circumstances, while our moral sentiments will be conditioned by backward-looking considerations, our actual treatment of the individuals who are the objects of these sentiments will be shaped (at least in part) by forward-looking considerations. This discrepancy between, on the one hand, the way we think and feel about one another and, on the other hand, the way we treat one another – if it is livable – will require a considerable loss of spontaneity. Moreover, on the basis of such a policy we may find ourselves treating individuals in inconsistent ways. For example, in certain circumstances utilitarian considerations may dictate that we punish some individuals for relatively minor failings while other individuals, whom we more strongly disapprove of, go unpunished (because punishment would serve no further purpose in their case). Similarly, we may find that two individuals who are the objects of similar negative moral sentiments should nevertheless be treated differently in respect of punishment because it will be of utility to punish one but not the other. Such circumstances may be unlikely or improbable, perhaps, but it is equally clear that they could well arise. In short, it may be argued that even if we concede to Hume that it is possible to curb our retributive practices in the way that he suggests, we will, nevertheless, have to pay a high price for such a policy in terms of hypocrisy, insincerity and inconsistency. We should be reluctant, the critic may argue, to allow such a large discrepancy between our thought and action, feeling and practice, to develop.

Second, the critic may continue to be unconvinced by the Humean account of the relation between retributive practices and the will. That is, it remains far from obvious that it is possible to disengage our (natural) retributive feelings and desires from retributive practices in the way that Hume's theory presupposes. It is possible, perhaps, for some (few) individuals to master their retributive sentiments and desires in such a way that they could abandon or suspend their retributive practices. (We may view such individuals as 'Spinozists' in practice, if not in thought.) Similarly, it is possible that all of us are, in some measure, capable of checking or curbing our retributive practices. Nevertheless, it is quite another thing to suggest that we are all (or even a sizable minority of us) capable of entirely suspending or abandoning our

retributive practices. Granted that our moral sentiments are inescapable features of human life, it is, surely, demanding too much of human nature to claim that we could – despite this burden – simply suspend or abandon our retributive practices if we so choose. Our moral sentiments and our retributive practices are too closely bound up with each other for that to be possible. Considerations of this general nature, it may be argued, suggest that the psychological foundations of Hume's complex theory of punishment are themselves suspect.

It seems clear, in light of the foregoing analysis, that both Hume and Smith encounter difficulties with their naturalistic approach to the problem of punishment. Moreover, as it stands, it is not evident that either strategy can overcome all of the difficulties which lie in their respective paths. In several respects, as I have indicated, Hume's theory is both more plausible and more satisfactory than Smith's (strong) positive retributivism. Nevertheless, whatever the merits of Hume's account, there remains a serious tension between the claim that we are naturally and inescapably committed to our moral sentiments and the supposition that we may suspend or abandon our retributive practices where and when they prove to be of no social utility. Moreover, it may also be argued that whatever the shortcomings in Smith's account, at least it does not license a radical discrepancy between our moral evaluations of people and the way that we treat them. Clearly, then, difficulties arise for Hume as well as Smith.

Those thinkers who wish to take a naturalistic approach to the problem of responsibility along the lines of Strawson should pause to consider the difficulties that Hume and Smith run into in this sphere. In particular, they should ask themselves where Strawson stands on these issues. Would he, for example, accept the Humean view that there exists a 'justificatory gap' between our moral sentiments and our retributive practices – or would he follow Smith in rejecting this claim? Similarly, would Strawson accept the Humean view that our retributive practices, unlike our moral sentiments, are *not* inescapable features of human life? All this remains unclear. Finally, as I have suggested, it may be argued that if it is indeed impossible to reconcile a naturalistic approach to responsibility with a credible and coherent theory of punishment then we have, on the face of it, strong grounds for rejecting such an approach. It is, accordingly, crucial that those who find merit in Strawson's discussion give further consideration to the rival strategies which Hume and Smith provide us with.

IV

My first objective in this paper has been to articulate and describe the nature and character of Hume's theory of punishment. There are, I

claim, two particularly notable features of Hume's discussion. First, Hume's theory of punishment rests on the foundations of a naturalistic theory of moral responsibility. Second, on the basis of this naturalistic foundation, Hume develops a 'mixed' or teleological retributivist account of punishment. These two elements of Hume's theory serve to distinguish his position from, on the one hand, those of Hobbes and Schlick, and on the other hand, from those of Smith and his followers. So interpreted, Hume's theory of punishment is, I have argued, not only distinctive, but also far richer and more complex than has hitherto been recognized. The historical significance of this conclusion is, I take it, quite self-evident.

My second objective has been to show that the contemporary interest of Hume's discussion lies primarily with his naturalistic approach to the problem of punishment. In particular, I have argued that Hume's discussion brings to light a number of interesting and important problems for Strawson's influential (and rather similar) naturalistic account of responsibility. In more general terms, Hume's attempt to interpret the implications of a naturalistic theory of responsibility for the problem of punishment, I maintain, sheds considerable light on the wider issue concerning the relationship between responsibility and punishment.⁴⁶ Within this general framework Hume's contribution is, I suggest, as illuminating as it is distinctive. Given that Hume's views on this subject have been largely overlooked, this is a conclusion of some consequence.⁴⁷

Received: July, 1989

Revised: April, 1990

46 This issue – along with others closely related to it – is discussed and analysed in an interesting fashion by Jeffrie Murphy and Jean Hampton in *Forgiveness and Mercy* (Cambridge: Cambridge University Press 1988). It is worth noting that Hume's complex theory of punishment provides an interesting point of contrast and criticism from which to judge and evaluate Murphy's and Hampton's (divergent) contributions to the contemporary debate. It may also be argued, from another perspective, that the interpretation and analysis of Hume's discussion provided here reveals the extent to which Hume anticipated – and perhaps indirectly shaped – some of the more interesting recent developments in this sphere.

47 For helpful suggestions, remarks, and/or discussions concerning various aspects of this paper (early drafts of which date back to 1983) I would like to thank Cora Diamond, Ross Harrison, John Marshall, Neil MacCormick, Bernard Williams and, particularly, D.D. Raphael and the editor and referees for the *CJP*: David Copp, Annette Baier and Marcia Baron. I would also like to thank audiences at Virginia, Chicago and Stanford for their comments on a paper of related interest. The final draft of this paper was prepared while I held a Mellon Fellowship at Stanford University.