Philosophy of Law
An introduction

2nd Edition

Mark Tebbit

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

12 Theories of punishment

The institution of state punishment is so widespread in the contemporary world that the question of justifying its very existence does not often arise. This is partly because the answer seems so obvious. Without any structure of positive sanctions in place, it is assumed, normal social transactions could not be governed by law. Without any mechanism for coercion or enforcement of legal norms, we could hardly speak of a legal system at all. Close analysis of the justification of punishment, however, reveals serious tensions, not only between competing theoretical perspectives but also between conflicting practical attitudes on questions about what kinds of punishment are morally acceptable. Various defences of the existing systems offer inconsistent accounts of the principles underlying punishment. They also collide with proposals for penal reform and even with frankly abolitionist arguments. What philosophical analysis over the last few centuries has aimed at is a clear examination of the principles in conflict here.

The problem of justification

We need first to remember why state punishment requires justification. Like many other common practices, punishment is at least *prima facie* problematic because it involves the kind of treatment that in any other context would be morally indefensible. As is well known, the history of the practice of punishment has spanned the full range, from the mildest of penalties to the most cruel and inhumane treatment imaginable. Even in an age when we like to think that most state punishment is relatively humane, policies such as the death penalty or imprisonment, involving the deprivation of life or liberty, clearly stand in need of justification.

It has often been found instructive to compare this question about justifying punishment with other, non-punitive instances of state coercion, which are also in their very nature morally problematic. Any state action or policy that involves the apparent violation of individual rights requires a special justification. Emergency measures in times of national crisis, the introduction of military conscription or national service, the Prevention of Terrorism Act and other suspensions of specific civil liberties are generally agreed to be

subject to special justification. Non-punitive coercion also includes practices such as the rationing of scarce resources, the requisitioning of property or internment of enemy citizens in wartime, compulsory purchase and the enforced quarantine of blameless victims of contagious diseases. None of these examples normally involve punishment or penalisation; nevertheless all of them are, to say the least, morally questionable and need to be justified.

Apart from the obvious consideration that the question of whether such measures can be justified will depend on the specific circumstances in each case, it at first seems that there is only one type of justification available, that of reference to ends and the invocation of 'necessary evil'. The greater the danger to be averted or the greater good to come out of the action — in terms of public safety, public health or national security – implies that the ends are desirable enough to outweigh the distress or inconvenience involved in the means. The assumption, then, is that the only justification available for such 'harsh measures' is based on an instrumental, means–ends type of reasoning.

Does it follow that the practice of state punishment is of the same order as these other questionable practices and that justification must follow the same instrumental route? Some philosophers have certainly thought so. Surely, it might be argued, the victims of such policies are blameless and, if such treatment can be justified, then the punishment of the guilty can *a fortiori* be justified along the same lines.

To show that this argument is too peremptory, we need to look more closely at the claim that all morally questionable acts can only be justified by instrumental reason. Is it possible to justify such an act without any reference at all to the good that will come out of it? One way to argue this would be to deny that the act was in fact morally questionable, or 'an evil' in itself. To argue that an activity is 'perfectly justified' is often to claim that there is nothing wrong with it, that those who disapprove are entirely mistaken. In a wider context, this is frequently said in defence of 'victimless' crimes, where no harm has been done. Another way, however, is to accept that in normal circumstances the act would be indefensible, but that in these unusual circumstances it is justified in itself as a response to another act. Thus, one might argue that 'a justified outburst' or 'justified anger' at an unwarranted attack on oneself or another is a morally appropriate response to provocation and contemplated without any view to the future. This could, of course, be represented as covertly instrumental, in that the justification is derived from the intention to defend one's interests, character or reputation. The ends again justify the means. This tendency, however, to reduce all justifications to forward-looking ones, with reference only to goals or objectives, obscures an important normative distinction.

To take the most serious example as an illustration of how this occurs, consider the procedures for justifying a military war against a clearly identifiable evil. For genuine justification here – as opposed to rationalisation before or after the fact – one has to present compelling reasons for taking

one course of action rather than another. To justify the resort to war, it is clear that the means employed must be a considerably lesser evil than the evil the war is intended to put a stop to. In addition, it needs to be shown that the response is appropriate and proportionate to the threat, that there are no effective alternatives to achieve the same objective, that non-combatant casualties are kept to a minimum, and that overall there is a realistic chance of success. Above all, though, the crucial condition for a just war is the antecedent one that the cause is just, that the war is genuinely against 'a clearly identifiable evil'.

While the instrumentalist strain of reasoning behind such justifications is predominant, this should not obscure the rival normative claim that the basic reason for morally endorsing a war precedes any instrumental thinking. Before any calculations of the prospect of success, or acceptance of the constraints aimed at minimising the negative consequences, the fundamental point is that it is already justified by the cause. For intrinsicalist thinking, the justification already exists; the course of action is justified by what has happened, such as an act of unprovoked aggression, not by predicted outcomes. As a chosen course of action, the war is claimed to be right in itself, not by virtue of events beyond itself. Having *the moral right* to fight a war is quite distinct from whether or not it is a good idea. Clear statements of war aims, though highly desirable, are quite distinct from an evaluation of the initial moral and legal legitimacy of the war. Entitlement precedes practicability.

Instrumental and intrinsic angles on justification do not necessarily conflict in practice, but there is a strong tendency in philosophical analysis for the partisans of each side to attempt to eliminate or marginalise the other, arguing about where the 'real' justification is to be found. The difficulty here lies in the fact that, while these two perspectives can frequently complement one another, they can and do easily come into direct conflict. This is sometimes represented as the clash between principle and pragmatism, or between idealised and realistic justification. This is not entirely appropriate, because it is easy to see how advocates of the two positions can swap roles in this respect, with the intrinsicalist prepared to justify anything for the sake of the just cause, and the instrumentalist with an eye to the future insisting on proper restraints in the conduct of the war.

Punishment justified by its effects

Something very similar to this kind of moral conflict and confusion lies behind the disputes over the justification of punishment. Given that state punishment can involve extremely harsh treatment, which in normal circumstances (which is to say, without good cause) would be regarded with horror, philosophers who adopt the instrumentalist outlook insist that the only justification for such drastic measures must lie in the compensating benefits that will come out of it. The standard practice of imposing long periods of

imprisonment for serious crime, for example, can only be justified if this practice as a whole is at least partly instrumental in the reduction or control of crime. If it cannot be justified with reference to these or some other social goals, then punishment inflicted for its own sake would be seen as nothing more than pointless cruelty, and hence manifestly unjustified. The basic requirement of justification is that the punishment has some definite purpose.

What this purpose might be varies enormously. To define it as the reduction of crime is the most general statement of the purpose of punishment, thus understood. For this reason, this perspective is often referred to as 'reductivism' (Walker 1991). This objective is in turn justified in terms of more general moral aims, such as the need for personal security and the opportunity to fulfil human potential. This in turn is justified until we reach a goal – such as the utilitarian standard of the general happiness – that does not need to be justified in any other terms. That is to say, we reach a goal that has non-instrumental value.

The reductivist aim is itself broken down into many specific penal strategies, with widely varying moral implications. The most obvious preventive strategy is the use of temporary or permanent confinement – either in prison or psychiatric hospital – to neutralise or incapacitate dangerous offenders. Consequently, the most basic justification of punishment is the need to remove offenders from society. Taken in isolation, this justification is morally equivalent to the enforced quarantine of carriers of deadly diseases. Second, one of the central reductivist strategies is the threat or the implementation of punitive sanctions in order to discourage or pre-empt the need for punishment. What this involves is the use of punishment as a deterrent, in a specific form to deter its recipient from reoffending, and in a general form to discourage others from committing the same kind of offence.

In addition to the goals of prevention and deterrence, reductivist strategies have included various attempts to adjust the behaviour of persistent offenders. The ideal aim here is to achieve the rehabilitation or reintegration of the offender into society. This rehabilitative ideal takes us beyond the scope of the standard consequentialist theories of punishment, primarily based on prevention and deterrence, because it ranges from the kind of moral education and reform strategies that see punishment itself as the agent of such effects, to approaches that regard the adjustment of the offender's behaviour as a preferable alternative to punishment. The latter includes the belief that offenders can be 'cured' of their criminal tendencies by various kinds of aversion therapy. Overall, though, this general goal of rehabilitation still exemplifies a broadly instrumental approach to the problem of justification, in that it rests on the assumption that this is the only way to justify either openly punitive practices or compulsory reformative alternatives.

The instrumental justification of punishment requires that at least one of these strategies lies behind its practice. The condition of its justification is that the evil of punishing is outweighed by one or more of these compensating benefits.

Justifying punishment retrospectively

For those who look exclusively to the present and the past for justification, the good that may come out of the practice of state punishment is contingent or incidental to the justification. Whether or not there exists a right to punish depends on what has already happened. The various social and individual benefits of the practice – the prevention and reduction of crime, the creation of greater peace and security, the effects on the offenders and their victims – are recognised as important considerations in themselves, but also as outcomes to be encouraged quite independently of the legitimacy of punishing. The right, in other words, precedes the good. The right to punish is derived from the offender's breaking of the law, or from his or her violation of the rights of others. How exactly the right to punish is derived from these past events rather than from projections into the future is, for the retrospectivist, the crux of the justification problem.

The main issue for retrospectivists, then, is how far they succeed in fixing a strictly non-consequential horizon. In its purest versions, the committing of a crime is both necessary and sufficient for the justification of punishment. The crime in some sense calls for or demands a punitive response. The classic statement of this claim was made by Hegel (1770–1831), who represented both crime and punishment in terms of the negation or annulment of its opposite. Punishment is justified because it nullifies or makes nothing of the crime; it negates it both legally and morally. This is why, morally speaking, it can be done, and why it has to be done; the state has both the right and the duty to punish.

This was one version of what is known as retributivism. Although the literal meaning of 'retribution' concerns the idea of the duty of repayment or reparation for wrongs committed, the connotations attaching to this ancient term are much wider. The core concept of the traditional theories of retribution is that of desert, indicating the principle that punishment should be given to people according to what they justly deserve, rather than to what we may feel is necessary for purposes of deterrence or rehabilitation.

Most of the other key retributive concepts revolve around this one. The right to retaliate with equivalent force against intentional violations of the moral code as enforced by the law is intrinsically linked with the idea that wrongdoing deserves punishment of the same level of seriousness as the offence. The general idea that retributive justice is the proper goal of punishment, rather than consequentialist calculations of outcomes, is also based on the idea of just deserts. Desert lies behind the retributive belief that the suffering inflicted by the punishment on the guilty is not an intrinsic evil to be regretted, but on the contrary a desirable state of affairs. This can only be so if those who have inflicted suffering on others can be said to deserve to suffer themselves. Each of these elements of the retributive justification is essentially backward-looking; the repayment, the retaliation, the deserved sufferings for past wrongs, all seem to be purely retrospective in their frame of reference.

Several other elements combine to produce the standard retributive justification. One of the more prominent is the Kantian insistence that the precondition of just punishment is that it treats offenders with respect for their dignity and capacity for free choice. Treating offenders as ends in themselves is the positive requirement here, but what is perhaps more important is what this prescription rules out, the use of others 'merely as a means' for greater social objectives.

Criticisms of the traditional theories

The most general and fundamental criticism of the forward-looking, instrumental approach is the claim that it is unjustifiably lenient in its implications for penal policy, that it releases the state from its obligation to punish crime in accordance with desert. Retributivists argue that, if desert is removed from the justification, nothing is left to prevent inappropriately light sentencing for serious crime. The role played by desert in this respect cannot be taken over by deterrence, because it is precisely the restriction of the goal of punishment to this unpredictable factor that destabilises the entire institution of state punishment.

It may at first seem paradoxical that the other main type of criticism focuses on the personal dignity and rights of those who are liable to be punished by consequentialists. The first aspect of this line of criticism is the belief that we – or the state – owe it to the offenders themselves to punish them as if they were free reasoning agents. If offenders are punished in accordance with factors extraneous to the fact and nature of their crime, or their degree of culpability, the implication is that they are being treated – contrary to the Kantian dictum – merely as a means to an end, rather than as responsible human agents. This idea that it is morally repugnant to regard people merely as objects for manipulation is applied generally to the practice of punishment, but it is thought especially relevant to the forcible re-educative and 'curative' rehabilitation programmes favoured by some consequentialists. A typically retributive view is that even the most severe punishment is less dehumanising than this kind of 'treatment'.

The second aspect of this line of criticism is that there is an irresistible logic whereby the unrestricted pursuit of utility leads to outright injustice. This is a criticism applied much more broadly to utilitarianism as a moral theory, but it becomes particularly focused in the philosophy of punishment. The danger in this respect is quite clear. If the administration of punishment, especially in sentencing, were to be guided solely by considerations of social policy and expediency, calculating the best probable consequences for society as a whole, there would be no limit to the potential injustice suffered by individuals or minority groups for the sake of the general good. There would be no restraint on favouritism or victimisation, or on the nature of the punishment. Even the complete absence of guilt would provide no good reason for not punishing, if the occasional scapegoating of the innocent

could be shown to be in the public interest. This, it is often said, is the logical conclusion of making the justification of punishment purely forward-looking. The overall criticism is that, when the core requirements of guilt and desert are removed, there remains no steadiness of purpose or stability in the philosophy of punishment. Every policy is subject to variation and experimentation.

It is in their attack on this inherent lack of control or limitation on the instrumentalist approach that retributivists score most heavily. Their insistence that prospective benefits – even the prospect of the complete elimination of a particularly threatening crime – do not give the state the right to punish, that this can only be derived from the prior fact and nature of the crime and what the criminal thereby deserves, puts them in a position to make the retributivist case persuasive by virtue of its focus on the injustice of punishing the innocent, which makes an almost universal appeal to moral intuition.

Weaknesses of retributivism

The weaknesses found in traditional retributivism fall under three headings: (1) doubts about its status as a moral justification; (2) problems with intelligibility; and (3) problems of rationality. Overall, it has been argued by critics that retributivism is morally dubious, that its key concepts are incorrigibly vague or ambiguous, and that it is based on feeling rather than reason. It will become clear that these sets of problems are interconnected, but it is important first to isolate them as distinct weaknesses.

The moral status problem

This is seen by many consequentialist critics as the basic and decisive one. Retributivism, it is said, does not really qualify as a moral theory of justification at all, because it is based on the premoral instincts to retaliate and take revenge. What this sanctions is the taking of sadistic pleasure in the self-righteous infliction of suffering. The point of constructing a modern moral theory is to civilise our instincts, not to give them free rein by institutionalising them.

Traditional retributivists hold that the state draws not only the right but also the duty to punish exclusively from the fact and nature of the crime, rather than from any benefits the punishment might produce. The duty as well as the right to punish is strictly backward-looking. This is the source of the most damaging criticisms of the retributive case. If the duty arises solely from what cannot be undone, it is argued, the state has no moral alternative but to exact retribution, regardless of the good or evil that may come out of it, even if a non-punitive response would be manifestly preferable. In this respect, it is argued, they are saddled with an outdated superstition, that every crime must be paid for, regardless of exonerating circumstances. On

this line of reasoning, the logical conclusion of retributivism is a completely irrational and vindictive insistence on punishment for its own sake.

The retaliatory character of the retributivist justification is central to Kant's philosophy of punishment and is rooted in the ancient biblical doctrine of *lex talionis*. A standard line of defence against the criticism that retributivism is rooted in an Old Testament morality of vengeance, from which we should long since have distanced ourselves, is that even in the original sources these maxims were intended to civilise rather than to urge vengeance. The idea is that the 'eye for an eye' maxim was a judicial rule aimed at limiting revenge to inflicting equivalent harm – that is, to take no more than an eye – rather than insisting that justice demands the taking of revenge. It has to be said that there is little support for this interpretation in the relevant passages, which do seem to insist upon responding punitively, rather than urging restraint in the response. It is quite possible, of course, that both meanings can be intended simultaneously. Even on the softer interpretation, however, the question of whether the returning of 'like for like' is really a moral response remains unresolved.

Although appeals to intuition are frequent in this debate, they do not seem to take us much further, because there is a deep intuitive conflict on this matter. While some find it self-evident that a serious offence intrinsically merits an equally serious punishment, others find it intuitively obvious that punishment for the sole purpose of retribution is pointless, that it serves no useful or civilised purpose. The retributive reply that the point of punishment is that it serves the purpose of criminal justice is met by the response that this is a particular, outdated and severe conception of justice that is inappropriate to a modern humane society.

Problems with intelligibility

In addition to the claim that the retributive attitude is morally reactionary, the second major weakness identified by critics is that all attempts to develop a systematic elucidation of its key concepts have led only to deeper mystification. What does it really mean, for example, to say that a criminal 'deserves' to suffer or be penalised for the crime? Does it mean any more than that we believe they ought to suffer for it? If so, how does this 'ought' mysteriously arise from the fact or nature of the crime? Why should it not be derived from, for example, the need to prevent and deter other such crimes?

The retributivist reply to this is complex. On the face of it, one might imagine, it is easy to argue that desert, like acknowledgement or gratitude, is an essentially retrospective concept and is intuitively intelligible as such. The difficulty that remains, however, is that of showing that it makes sense to say that either a moral right or an obligation to punish can be generated by the simple reflection that we commonly use this retrospective concept.

More specifically, what is the reasoning behind the retributive claim that the punishment should 'fit' or 'match' the crime? In the context of the death

sentence for murder – 'a life for a life' – it is conceded by critics that this is at least intelligible; however, what would the matching sentence be for fraud, for treason, for rape? The idea of even a rough correspondence in this sense between a crime and its punishment is said to be of little practical value. Ultimately, it is argued, the retributive language of desert is mere rhetoric to mask the absence of an intelligible justification.

The standard retributivist reply to this criticism as a whole is that the intelligibility of desert as a justification is exhibited in the principle of proportionality, the violation of which – handing out disproportionate punishment – is a clear injustice. To punish a minor theft, for example, more severely than an armed robbery is manifestly unjust. Starting from such examples, it is easy to construct parallel scales of seriousness in criminal offences, on the one hand, and penalties or sentences, on the other. That one should match the other is a requirement of desert, not of any consequentialist calculation.

There are basically two problems with this. The first is that any ranking of criminal offences according to desert is controversial. Which is the more serious, meriting more serious punishment: robbery with violence, or a non-violent crime such as fraud, the illicit proceeds of which are much greater? Second, even if a scale of desert is settled, this only solves the problem of the relative severity of the punishment. It does not fix the level of the mean, which in practice is relative to the standards of a particular society. Nor does it fix the range (minimum and maximum) or the spread (ratio of one offence to another) of the scale. A retributivist might well reply that the very recognition of these problems amounts to a tacit admission that desert is the appropriate basis for justification. The consequentialist can in turn reply that the principle of proportionality gives us at most a justification for treating some offences as more serious than others; the scale might as easily be applied to the appropriate degree of disapproval or reprimand, as to the institution of punishment. What this demonstrates, however, is that the justification is difficult, not that the idea of desert as justification is unintelligible.

The most notoriously vague aspect of the retributivist account originates in Hegel's theory that the punishment constitutes an annulment of the crime. The idea at the heart of this theory is that the act of punishment constitutes a denial of the legitimacy of the criminal act, in Hegel's dialectical terms, a rightful negation of the criminal negation, leading to a moral reaffirmation of the legitimacy of the rightful order. Although this perhaps comes the closest to a purely retributive justification, problems of intelligibility have frequently been voiced. How can punitive action nullify or 'make nothing of' an offence in which death or permanent injury has been caused? How can the situation prior to the crime be restored? These objections, however, rest upon a misinterpretation of the sense of negation intended by Hegel. It is not the offence itself that is annulled or 'cancelled out', it is the implicit claim to a morally and legally invalid legitimacy that is 'made nothing of'. For Hegel,

the dialectical process of crime and punishment is a struggle for the recognition of the validity of the moral order that has been challenged. The justification of punishment is purely retrospective because it is drawn solely from the challenge to the moral and legal authority of the state.

Problems of rationality

In addition to doubts about moral soundness and intelligibility, the third type of weakness commonly attributed to the retributive approaches is the absence of a rational basis for the justification of punishment. The problem here is that, even if we can make it clear what is required for a retributive justification and give an intelligible account of the state's right and duty to punish without reference to consequences, we cannot give any good reasons for preferring this justification, without such reference. The best that can be done is to clarify by example the intuition that the guilty deserve 'to be brought to justice'. The problem here is that most phrases that seem to capture the elusive meaning of the retributive justification are highly emotive ones. The general charge in terms of rationality is that retributivism is entirely dependent on intuition and the negative emotions of anger, resentment and hatred. An important accompanying criticism is that attempts at elucidation depend too much on symbol and metaphor, not enough on reasoned argument.

There are essentially two ways in which retributivists can reply to this. They can either accept that retributivism is not rational, but argue that it is nevertheless morally defensible, or they can reject the instrumentalist model of rationality and argue that a rational account of retributivism can be given. With the first approach, it can be argued that the retributive emotions, properly controlled and channelled by law, express a wholly legitimate response to crime and that this in itself constitutes a justification. The reason for preferring the retributive justification is that it is held to be psychologically and morally realistic, in that it conforms with the sentiments of disapproval or abhorrence that most people feel in response to serious crime. This approach questions the assumption that justification as such has to be rational (Mackie 1985: ch. 15).

The second way is more difficult. It is insufficient – though quite correct – to point out that retributivists do not lack reasoned arguments; the question is about the basis of these arguments. What has to be shown here is that a strictly non-consequential justification can sensibly be called a rational one. If 'rationality' simply means 'thinking in terms of consequences', any non-consequential thinking becomes irrational by definition. If, on the other hand, its meaning is wider than this, including retrospectivist thinking in terms of pre-existent rights and the duty to respect them for their own sake, then the constraints on the application of the narrower rationality will themselves be seen as an integral part of rationality, rather than as emotive constraints.

Modifications and compromises

It should be clear from the discussion to this point why, unless one side gives way on an important point of principle, the two positions cannot be simply combined into a unified theory of punishment. There have nevertheless been numerous attempts, especially in the second half of the twentieth century, to modify the theories in such a way that they might complement one another. The idea behind the 'mixed theory' is to search for a common ground on which some of the opposing elements can be synthesised into a coherent theory that will either reflect the reality of existing legal practices or provide the basis for realistic proposals to reform the current system. From this standpoint, neither the instrumental justification nor the retributive one in their pure and intransigent versions are seen as realistic in either sense.

There are three important possible structures for the mixed theory to adopt. It can be (1) retributive in its basic justification, making concessions to the demands of social policy; (2) instrumentalist in its basic justification, making concessions to one or more of the retributive principles; or (3) more radically innovative in that the basis of justification is extended across both areas, so that both retributive justice and social value are necessary conditions, but neither alone are sufficient.

Strong and weak retributivism

The first option has been a very popular one. What the moderation of retributivism involves is the distinction between a maximum version, insisting on both the right and the duty of the state to punish, and a minimum version that relinquishes the duty, insisting only on the prior right to punish and the forfeiture by the criminal of the right not to be punished. The main advantage of this is that it avoids the range of criticisms relating to its moral status. On the minimum interpretation, the state only exercises its right when there actually is a non-retributive point, which is to say that it takes a flexible approach to prosecution and sentencing, allowing that consequential considerations can override the *prima facie* duty to punish.

The maximum interpretation is clearly indicated by the traditional versions of retributivism. Kant's insistence on the solemn duty to execute every last murderer (1887: 194–201) is the paradigm case of strong retributivism. Hegel's theory of annulment also implies the inseparability of the duty and the right to punish; the theory is an explanation of why the state is morally obliged as well as entitled to invalidate the crime. It cannot let the offence stand. The case for minimalism, however, has been defended by modern retributivists such as Ross (1930), Armstrong (1961) and Mundle (1954).

There is certainly some plausibility in arguing that the minimal thesis is more in line with actual practices; the commuting of sentences and reprieves, the royal prerogative of mercy, the powers of the Home Secretary, judicial discretion and mitigation of severity of sentence have played a prominent role in the history of the English legal system, and also many prosecutions

are not held to be in the public interest. In short, the right to punish is not always exercised. Against this, however, some sentences have been mandatory and many offences are not regarded as subject to discretion.

The relevant question here, however, is about the kind of principles in operation. Flexibility and mitigation of sentence tend to be desert-based – focusing on degrees of responsibility – rather than consequentialist. To the extent that this is true, the mixed theory is not a concession to consequentialism at all. If, on the other hand, the theory were to accept the principle that consequentialist considerations should govern sentencing as a proposal for systematic reform, it would be making so many radical concessions that it would be difficult to see it as retaining any more than a formal commitment to retributivism. Either way, the minimal version, abandoning the duty to punish, does not seem to provide a basis for a real compromise. The dilemma we are left with is that, while the Kantian strong version is too strong, the weak version either makes no real concessions or virtually dissolves as a retributive theory.

Lex talionis *and unfair advantage*

A different kind of attempt to modify the retaliatory character of retributivism is represented by theories that shift the justification from the doctrine of *lex talionis* to the idea that desert is based, not on the right of the state to retaliate against offenders, but on the right and duty to remove the advantage unfairly gained by the offender's refusal to play by the rules. The duty is towards those who have not taken similar advantage, those on behalf of whom the state acts, and the focus is on the injustice towards the law-abiding. Offenders deserve to suffer in a measure equivalent to their offence, not by virtue of an ancient moral law commanding such equivalence, but because the failure to cancel the advantage is an injustice.

This line of thought is aimed at defusing the criticism that retributivism is based solely on the vengeful emotions. The idea that the suppression of unfair advantage is morally mandatory is aimed, not at changing the substance of the theory, but at providing it with rational rather than emotional backing and thereby making it more intelligible.

Consequentialist compromises

There are a number of important theories that are explicitly instrumentalist in their basic assumptions, but that are designed specifically for the purpose of reconciling this kind of justification with retributivism. The two most influential were developed by Rawls (1955) and H.L.A. Hart in 1959 (Hart 1968).

The essential feature common to each of these theories was the claim that the problem of justification in punishment theory cannot be expressed by one question – such as 'How is punishment justified?' – but must address a more discriminating set of questions, the answers to which are different in

kind. It is the difference between these answers that is supposed to create the ground for a compromise, or for a combined theory of punishment.

Rawls's opening distinction is between (1) any practice or system of rules, such as a game, a governing assembly or the institution of punishment, and (2) a particular action falling under these rules, such as a move in a game, a parliamentary enactment or a judicial decision, any one of which is governed by the relevant system of rules. Justifying a practice, Rawls argued, is quite different from justifying any of its instances. With the practice of punishment, what we are determining is the initial purpose of setting up the institution and punishing anybody at all. This, he claimed, can only be justified in utilitarian terms, as furthering in some way the interests of society. Nobody, he believed, would want to argue that the very *purpose* of punishment was to match wrongdoing with suffering. With a particular instance of punishment, by contrast, the conviction and sentencing of an individual lawbreaker can only be justified in terms of that individual's guilt, or the fact that he has broken the rules of the practice.

What Rawls was arguing was that it would be inappropriate to try to justify any such punitive action in forward-looking, consequentialist terms. While the legislator laying down the law looks to the future, the judge applying the law looks to the past. In this way, Rawls argued that utilitarians and retributivists both have a legitimate point, and that the two perspectives can be combined by recognising this distinction.

Hart's starting point is similar to that of Rawls, in that he distinguishes between a 'general justifying aim' behind punishment as a whole, and the specific principles of justice guiding and restricting the application of punishment. He then rejects retribution as the general aim and argues that it is legitimately to be found in the principles constraining the operation of utility. This he describes as 'retribution in distribution', which he finds morally defensible. The main feature of this distributive justice is the requirement that guilt is a necessary condition of punishment. Punishment as a whole is justified in the first place – as it had been in Rawls's version – by the general justifying aim of its beneficial consequences, primarily crime reduction. What Hart was seeking, in his own words, was 'the middle way between a purely forward-looking scheme of social hygiene and theories which treat retribution as a general justifying aim' (1968: 233). One of Hart's central themes was the need to untangle the conceptual confusion caused by utilitarians and retributivists alike in failing to see the distinction between the general justifying aim and the principled constraints upon it.

Both Rawls and Hart worked on the assumption that, if the implications of these distinctions could be made clear, only the most intransigent of old-fashioned retributivists could fail to see that the justifying aim must be consequentialist, especially given the recognition of the real place of the retributivist principles in the practice of punishment. The implication that they both attached great importance to, that the problem of punishing the innocent could be avoided by combining the utilitarian general aim with

'retribution in distribution', was thought to be decisive. That these assumptions were unduly optimistic was to be clearly demonstrated by the retributivist revival in the 1970s. The main practical concern in this revival was to reinstate the priority of desert over deterrence as the fundamental justification.

Punishment as communication

The theme of expression and communication implicitly runs through all the literature on punishment. Even where it is not explicit, the idea of a negative value judgement embodied in or accompanying the punishment is present in any attempt to justify it. In recent times, the close attention given to the communicative aspect of punishment as its inner meaning has been developed in order to find the source of the conflict between the different perspectives on justification.

Communicative theories have been predominantly but not exclusively retributivist. The idea that the inherent meaning of punishment is to convey a message of emphatic denunciation, whether to the offender, the victim of the crime, or the society in whose name the state punishes, can as easily be interpreted as having an instrumental function (as serving the ends of crime prevention or moral education) as being an end in itself. This kind of theory has nevertheless been associated mainly with the retributivist revival since the early 1970s, with attempts to clarify, reinforce or amend the traditional versions of retributivism. The commitment of many of the new retributivists to theories of communication should be seen partly as an attempt to move the debate forward, partly as a continuation of the old project of defending the moral respectability and rational intelligibility of punishment as retribution.

Robert Nozick: connecting with correct values

Robert Nozick's retributive theory is an attempt to do both, to illuminate the old tradition within the framework of a distinctive communicative theory. For Nozick, 'retributive punishment is an act of communicative behaviour' that communicates the unwelcome message to the offender, 'this is how wrong what you did was' (Nozick 1981: 370).

Taking as a guiding structure the simple formula $r \times H$ (degree of responsibility, multiplied by actual harm done), Nozick argues that retributive punishment is justified to the extent that a wrongdoer intentionally causes harm. The retributivist claim that the intentional committing of the crime is sufficient justification is expressed by Nozick in his central argument that the unwelcome message of wrongness forcefully delivered by punishment constitutes a reconnection with correct values of those who have flouted them through criminally harmful acts, thereby disconnecting themselves from these values.

Recognising the retributivist need to detach the justification from any hint of instrumental objectives, Nozick contrasts his own version with what he calls teleological retributivism, which looks primarily for a positive effect on and a response from the offender. Despite the forward-looking aspect and the reformative implications, this is still retributive because it is a message of condemnation merited by the wrongdoing. On his own version of reconnection, however, the retribution is justified even if the message is entirely unsuccessful. All that is required is that the message is sent and received. In cases where the message has no effect, the connection is still imposed via punishment. 'The act of retributive punishment itself effects this connection' (Nozick 1981: 374). To be punished is to be connected. Like most previous retributivists, Nozick regards a positive response by the offender to the message as a valuable bonus, but unlike them he describes this as an *intensification* of what is already achieved without it.

The difficulty many have found with this lies in the highly teleological flavour of the phrase 'reconnection with correct values', which sounds as though it must depend on a project of moral improvement. Despite his denials that any reformative goals, successful or otherwise, are necessary for this justification, Nozick's purely non-teleological retributivism remains difficult to interpret.

What he seems to mean by this is that, right from the outset, the imposition of punishment is an involuntary connection with correct values of those who have broken and resisted this connection, thus 'flouting' correct values. This enforced connection is already justified solely in terms of connecting even if the offender subsequently comes nowhere near to understanding or responding to the message and is determined to reoffend as soon as released. If the offender does show signs of internalising the correct values, exhibiting understanding and remorse, what is happening is not something qualitatively different, it is the connection – already in place – beginning to work.

One thing Nozick is trying to defuse is the standard criticism that punishment without beneficial results is pointless. What he is implicitly denying in his insistence on the continuity of the communicative behaviour of the punisher is the radical breach between intrinsic and instrumental value and justification. As he sees it, the latter might or might not emerge from the former.

Jean Hampton: defeating the wrongdoer

The retributive emphasis on rights is of particular importance to the attempts to interpret the tradition in the light of the idea of punishment as communication. The pressure from instrumentalist criticism is resisted by viewing retributive punishment as the legitimate sending of a message that responds to criminal violations or infringements of the rights of others. The justification originates in the criminally coercive invasion of rights, rather than in the instrumental value of the denunciatory message.

Jean Hampton places the issue of rights violation at the centre of her communicative theory. Following up Nozick's theory of connection or 'linkage' to correct values, she is less concerned with detaching the justification from moral reform or other penal policies than with explaining punishment as an inherently justified response to the rights-violating message sent out by the offender, and to explain why the state is at fault if it fails to respond with the appropriate message. Her essays rest on Kantian ideas about respect and human value; they offer a modern interpretation of the *lex talionis* and of the Hegelian theory of annulment (Murphy and Hampton 1988).

Hampton's central idea is that of 'inflicting a defeat' on the wrongdoer, whose violation of the rights of others is seen as sending out an objectively demeaning message of domination or mastery over the victim. The reason tough punishment, inflicting suffering in equal measure to the seriousness of the violation, is essential to the communication of the message is that the defeat must be a real one, rather than a merely symbolic denunciation. When society fails to deliver the appropriate punitive action, we allow ourselves as a whole, and the victim in particular, to be demeaned and defeated.

The message conveyed by the wrongdoer who invades the rights of another is that the other is of less personal worth than he or she actually is. At the same time as it expresses a diminishment of the worth of the other, the wrongdoer's message lays false claim to his or her own superior worth. It is the rightful correction of this false representation of their relative worth that justifies the punishment. The purpose of retributive punishment is to bring down the exaggerated claims about the offender's worth, while at the same time, through the same act, reasserting and restoring the worth of the victim. If the state fails to administer punishment sufficiently serious to bring about this defeat, it implicitly endorses the false representation of the relative worth of the rights of invader and victim.

Hampton calls this representation of relative worth 'false evidence of mastery'. It is this evidence, she argues, that is the proper object of Hegelian annulment. The negation of the evidence for a distorted representation of the relative worth of offender and victim restores the true picture in the same way that a scientific disproof of misleading evidence uncovers the true picture of a natural reality. It does not make the evidence disappear; it simply discredits it and shows that it does not prove what it appears to prove. This 'refutation' can only be achieved through punishment, because this is the only practical demonstration of the invalidity of the evidence and vindication of the victim's true worth. To punish is to force the would-be master to suffer defeat at the hands of the victim or the victim's agent. No amount of verbal or symbolic denunciation can inflict this defeat.

The point of Kant's insistence that the failure to punish every murderer implicates the entire community in the crimes left unpunished is explained by Hampton as a general insight into the significance of the message conveyed by punishment. In terms of her theory of false evidence, she argues that, when we abandon the retributive goal of punishment, we condone the false

evidence for the inferior worth of the victim, because it stands undefeated. In allowing it to stand, we 'acquiesce in the message it sent about the victim's inferiority'.

Her reinterpretation of *lex talionis* is that this is a formula demanding that the wrongdoer suffer a defeat on a similar scale to the one he or she has inflicted. Hampton argues that this should be restricted to the principle of proportionality, whereby the greater the offence the greater the defeat needed to reassert the victim's value. What she claims can be removed from the traditional doctrine is the severity of the 'eye for an eye' demand for equivalence, for the punishment to match the crime. The retributive punishment designed to nullify the demeaning message can be constrained by upper limits determined by the Kantian demand for the respect for the humanity of the wrongdoer, even when the wrongdoer has not shown similar respect for the humanity of others. She argues that this demand for respect imposes a ceiling on the severity of punishment, the function of which is to defeat without dehumanising and degrading. By rooting her demand for these humane upper limits to punishment within her own theoretical framework, rather than in *ad hoc* consequential reasoning, she seeks to avoid the old criticisms that retributivism is by its own logic necessarily barbaric.

Hampton's account as a whole certainly throws light on how the Kantian and Hegelian theories can be plausibly interpreted in line with contemporary penal policies, and it has radical implications both for minimum and maximum sentences. The idea that only punishment and the actual 'lowering' of the offender can deliver a real defeat of the criminal offence is perhaps more convincing than earlier retributivist accounts; it probably comes closer than most theories to bridging the gap between the rational case for denunciation of wrongdoing and the case for hard penal treatment.

There are, however, several serious problems. It is dubious that the restriction of *lex talionis* to proportional punishment, relinquishing the idea of it 'matching the crime', counts as a retaliatory principle at all. It is not even clear why it is distinctively retributive, as we shall see in the next section. Second, there is a stronger than usual emphasis here on personal worth and the lowering of personal value. This kind of account is at its most convincing in the context of offences against the person. When the emphasis is on crimes with victims who suffer irreversible physical or psychological damage, it is easier to find this justification plausible. It is more difficult to extend it to the full range of crime. Is it really plausible to interpret ordinary offences of theft as attempts by criminals to elevate themselves above their real moral status and to demean their victims? How does it fit impersonal crime, in which the victims are anonymous? The unfair advantage explanation seems more appropriate to such cases.

Instrumentalist censure and reprobation

For instrumentalists, the important question about the message communicated by punishment is how this expression of disapproval or condemnation

can be most effectively used. Thus, Walker and Padfield (1996: 116) describe it as one sentencing strategy among others, aimed specifically at increasing disapproval of the offence and increasing respect for the law. Braithwaite and Pettit (1990: 160–4) regard the censure or reprobation of the offender as an integral part of punishment, but only for the purpose of what it can achieve in terms of reducing crime and protecting people's freedom. They reject what they term the 'intrinsic reprobationism' of Nozick, von Hirsch and others, all of whom promote the idea that reprobation is a good in itself. In particular, they argue that the need for effective denunciation or reprobation does not support the retributive insistence on some degree of hard treatment, on the grounds that this can only be justified as a last resort, when prevention and public safety demand it.

Desert and deterrence in sentencing

One of the intractable difficulties that seems to have defeated the retributivist imagination is the problem of devising a method to determine the general level of what is deserved. It is much easier to find consensus on the question of relative desert, on which kinds of offender and offence deserve greater or less punishment than others, than it is to set the average level, reflecting the initial equivalence between the punishment and the crime. How do we determine, for example, whether the respective punishments for burglary and armed robbery should be five years and ten years imprisonment, or five months and ten months? This 'absolute' question is sometimes obscured by the arguments for proportionality in sentencing.

More recent theories of desert have focused on the question of how deterrence-based sentencing stands in relation to the desert requirements in terms of both equivalence and proportionality. Given that deterrence is the central factor in the attempts to control or reduce crime, the crucial question concerns the compatibility between this goal and the pursuit of justice in the sense of adhering to both of these principles. It should be noted that the orthodox judicial opinion in Britain and the USA today is that this requires a delicate balance between rights and social policy, and that this balance is largely achieved.

For some theorists, it is a question of which way the balance should be tilted, towards maximising crime reduction or safeguarding justice. For others, this balance is wholly fictitious, concealing the reality that in many areas of crime hard decisions have to be made, seriously compromising either the reductivist strategy or the demands of justice. Alan Goldman (in Simmons *et al.* 1995) has argued that a paradox or dilemma presents itself as soon as we compare the requirements of deterrence and desert. Given that the rate of apprehension and conviction for any type of crime is always less than perfect and often very low, those who in fact are caught and punished have to pay the price for the low rate of conviction, by serving sentences that are always greater than the crime would otherwise merit in

terms of the principle of equivalence. The paradox is that, while the resulting injustice, which is tantamount to punishing the innocent, is quite intolerable, the relaxation of this policy, allowing a massive increase in the criminal violations of the rights of innocent victims, would be equally intolerable. Goldman's conclusion is the unremarkable one that the only appropriate response is to increase levels of detection and attack the socio-economic roots of crime.

He deals with three possible objections to this account of sentencing practices. The 'exceptive clause' argument, that the excessive punishments are justified by the harms greatly outweighing the rights, is rejected on the grounds that there is nothing exceptional about them. The 'fair warning' objection, that if excessively harsh penalties are well-advertised, then the criminal has only himself or herself to blame for ignoring the warning, is rejected on the grounds that this would justify any degree of punishment, however extreme, which it clearly would not. The 'lynch mob' objection, that the criminals are themselves being protected from angry public reaction by excessive punishment, is rejected on the grounds that such reactions would justify punishing the vigilantes rather than the criminals.

Given the premises of the argument, these replies sound persuasive. The assumptions Goldman is making, however, are certainly questionable. We have to ask whether effective deterrence in general really requires more punishment than is deserved. The more usual retributive objection to deterrence is not that it is inherently excessive but that it has a destabilising effect on justice, in that it fluctuates according to current social circumstances and type of crime rather than according to equitable criteria. The retributivist conclusion is that lower sentences than desert demands are as common as sentences in excess of desert. With unusual crimes that are unlikely to be emulated, the need for deterrence is negligible, but the retributivist still insists that there must be an equivalence between harm intentionally caused and the punishment. With exceptionally long sentences, such as the 'life tariff', it is quite plain that desert outlives the need for deterrence. What Goldman's argument assumes is that the general level of desert is actually very low, relative to current sentencing practices, such that it will always be less than what is needed for deterrence. The mistake behind this is the assumption that the imperfect rate of detection and conviction is the only factor governing the operation of deterrence. It is, however, often applied quite independently of such considerations, for example to discourage the spread of a new kind of fraud, which would have nothing to do with making people suffer additionally for those who have not been caught.

Another argument that works against the Goldman paradox is found in the influential 'desert-band' theory put forward by Norval Morris (Duff and Garland 1994). This takes the edge off the equivalence problem by introducing flexibility into the determination of how much punishment any particular crime merits. If the attempt to determine the level of punishment deserved is not burdened with the assumption that there must be a precise

equivalent, a range can be established, within which the punishment is morally acceptable in terms of desert. Desert is seen as a limiting principle, setting 'the outer limits of leniency and severity which should not be exceeded'. Thus one might argue that, for example, a particular kind of serious assault merits at least six months, at most two years imprisonment, rather than fixing it at exactly two hundred days. With a rigid equivalence principle such as Hegel's, one day too many or too few would be an injustice. Morris's intention is to promote a sentencing practice whereby there is room for manœuvre or 'fine tuning' within this desert-band, applying lower or higher sentences for purposes of incapacitation and deterrence. This, for Morris, is the model upon which discretionary sentencing should operate.

How does this work against the Goldman paradox? The overall point of Morris's argument is that retributive justice does not have to clash with the pragmatic goals of crime reduction, which can be accommodated within the desert-band. It means that increasing the sentence within the minimum–maximum range solely for purposes of deterrence does not constitute an injustice. It does not mean that the tension is eliminated, because there is still an upper cut-off point beyond which deterrence strategies should not go, and a basic minimum that should not be ignored. Thoroughgoing consequentialists will still reject this as an obstacle to rational social policy. It has the potential, however, to reduce the tension between desert and deterrence.

Morris's desert-band theory has greater appeal, in that it offers both an explanation and a kind of justification of common sentencing practices, but it faces serious objections, the most obvious of which is that it merely shifts the problem of determining equivalence from an exact point to a range. Why, for example, should the range itself be fixed at two to four years, rather than four to eight? Why, indeed, should it be a prison term at all? Why not a ten- to twenty-minute severe reprimand?

Even if one accepts that the substitution of a range for a precise point facilitates the intuitive setting of equivalence, Morris's theory faces other objections from a retributivist position. There is, for example, nothing to prevent the range of deserved punishment from being manipulated and overstretched to accommodate grossly unjust sentences. It also explicitly sanctions inequality in sentencing and violates the principle of proportionality. Morris defends his rejection of the principle that 'like cases should be treated alike' on the grounds that equality in law is at most a guiding principle, to be balanced against other values and suspended whenever outweighed.

This was exactly the point that was rejected by von Hirsch (1993) in his defence of the 'commensurate desert' principle, which links the idea of just deserts with the principle of proportionality and 'commonsense notions of equity', ruling out disparate sentences for identical offences. Norval Morris's desert-band theory is unacceptable, he argues, because it holds the principle of equal treatment to embody merely one value among others, to be

dispensed with whenever desirable on utilitarian grounds. Von Hirsch, by contrast, insists that there should be a presumption in favour of the principle of commensurate deserts, that it should have *prima facie* controlling effect on sentencing, only giving way in exceptional circumstances.

Conclusion

The intention throughout this examination of the justifications has been to dispel some of the misconceptions surrounding the debates on punishment, in particular the one that sees two well-defined camps in stark opposition. At the same time, I have tried to show that there are nevertheless two very different modes of moral thinking about punishment, as there is with war and other instances of serious state coercion. Any well-informed reflection on the theory and practice of sentencing reveals a permanent tension between these two ways of thinking.

The search for a coherent link between these two approaches will certainly continue, but what most contemporary philosophers are anxious to avoid is a repetition of past mistakes, the most obvious of which is a simplistic endorsement of the uneasy compromise embodied in the contemporary institutions of punishment, combining the goal of controlling and reducing crime with retributive 'elements' in sentencing. The search continues, not so much for a 'mixed theory' that tries to establish a compromise between incompatibles, but rather for the articulation of a unified theory that combines the two sides, if not into a single position, at least into a framework in which genuine dialogue is possible. What is sought is a realistic theory that tries to root the justification in the kind of punishment which is the least ineffective, holding out some prospect of actually working towards the achievement of its basic goals, which in themselves neither exclude the proper concern with blame and desert, nor set up the kind of institutional framework that will systematically violate any of the principles of justice highlighted by the retributivists.

Study questions

General question: Can punishment be justified? If so, how?

Further study questions: Compare the merits and defects of forward-looking and backward-looking justifications of punishment. Is it possible to synthesise them into a single theory of punishment? What place should the concept of desert have in a theory of punishment? Might it be possible to base a theory of punishment solely on deterrence? Should levels of sentencing be fixed according to desert or deterrence? Does Nozick's communicative theory make better or worse sense of retribution? Is the Hegelian theory that punishment annuls the crime, intelligible? Evaluate Hampton's version of annulment theory.

Suggestions for further reading

Outstanding general books on the philosophy of punishment include Honderich (1976), Hart (1968), Lacey (1988) and Ten (1987). More introductory accounts can be found in Lyons (1984: ch. 5), Harris (1997: ch. 5) and Murphy and Coleman (1990: ch. 3).

Very useful anthologies of influential traditional writings and modern essays include Grupp (1971), Acton (1969) and Duff (1993). Two anthologies are particularly useful for the more recent developments in this area, the Duff and Garland reader (1994) and the *Philosophy and Public Affairs Reader* (Simmons *et al.* 1995).

Any attempt to attain a comprehensive view of twentieth-century and contemporary philosophy of punishment should begin with the hard core of influential arguments contained in these anthologies. The key passages from the classics are in Kant (1887: 194–201), Hegel (1942: 68–73) and Bentham (1970: 158–73).

In the early to mid-twentieth century, the most significant writings were those of the mixed theorists Ewing (1929), Rawls (1955) and Hart (1968); and the defences of retributivism by Mabbott, Mundle and Armstrong (Acton 1969).

As a representative sample of late twentieth-century writings, one should read Murphy (1979, 1987), Andenaes (1974), Kleinig (1973), Cottingham (1979), Mackie (1985: ch. 15), Primoratz (1989), Nozick (1981: ch. 4), Hampton (1984), Murphy and Hampton (1988), Walker (1991), Braithwaite and Pettit (1990) and von Hirsch (1993). On particular problems relating to sentencing, see Walker and Padfield (1996), Gross and von Hirsch (1981) and Duff (1993: part IV).