

2

Justifying Punishment

Most people think that punishment of wrongdoers by the state can sometimes be justified. We may dislike punishment, but we think that it is justifiable to do it to some people. That is so even though punishment involves the intentional infliction of harm or suffering on others. Furthermore, many people think that the state has a duty to punish offenders; that it has breached the duty it owes to its citizens when it fails to punish at all or it fails to punish enough.

One basic question to ask about punishment is: what is, or at least can be, valuable or good about it? To justify punishment, given the financial cost of punishment and all of the harm that it causes, we must point to some good-making feature that it has—something sufficiently important to justify the costs of punishment and the harms that it inevitably causes. The urgency of this question is intensified in virtue of the fact that punishment involves the intentional infliction of suffering. Making people suffer is normally bad, and making them suffer intentionally is doubly bad. What could possibly justify the systematic intentional infliction of suffering on other human beings?

One possibility is that the infliction of harm on offenders is intrinsically valuable—that is, valuable in itself. Another possibility is that it is valuable instrumentally—that is, valuable in virtue of the further effects that we bring about through the infliction of harm. A third possibility is that it is valuable in both these ways. Understanding what might be good about punishing a person is one essential step in the justification of punishment. It provides the reason or reasons that we could give in favour of punishment.

The question on which we have just been focusing, the question of what is valuable about punishment, concerns the *immediate* justification of punishment. If we punish people, and we are asked to justify ourselves, what we take to be the immediate justification of punishment will be the first answer that we will offer. Some people will say that punishment is good because it is intrinsically good that offenders suffer. Others will say that punishment is good because of its deterrence effects. Still others will say that punishment is both intrinsically and instrumentally good. In this book I will reject the idea that the punishment of wrongdoers is intrinsically good. If we are to justify punishment, it is to its instrumental value that we must appeal.

The immediate justification of punishment must be defended in the light of deeper issues in moral and political philosophy. When we reflect on punishment we naturally engage these deeper issues, and this is one reason why punishment has

attracted so much philosophical attention. We are asked to embed the justification of punishment in a more comprehensive moral and political philosophy. Why? Because different comprehensive moral and political views give rise to radically different views about the permissibility of harming others.

Very broadly speaking, we can distinguish between consequentialist and non-consequentialist views of morality. On a consequentialist view, whether it is right or wrong to harm a person is to be determined solely by the consequences that harming that person would have. As we will see, this basic consequentialist idea admits many variations. Consequentialism can be contrasted with a range of non-consequentialist views. Non-consequentialists believe that the consequences of harming others are not all that matters in determining whether harming them is permissible. As non-consequentialism is defined negatively, it is unsurprising that there are also many non-consequentialist views. For there are many different ways in which we might depart from the consequentialist claim.

A prevalent idea in the philosophy of punishment is that there is a natural relationship between the two immediate justifications of punishment that I offered and these two comprehensive views of morality. It is often assumed that those who think that punishment is intrinsically valuable—those who think that it is good in itself that offenders suffer—will be non-consequentialists. Those who think that punishment is instrumentally valuable—those who think that it is to be justified in virtue of its deterrence effects—will be consequentialists. The main aim of this chapter is to show that this popular idea is fundamentally mistaken. The debate about whether punishment is intrinsically valuable or whether it is instrumentally valuable has almost nothing at all to do with the debate about whether some version of consequentialism or non-consequentialism is the best comprehensive view to endorse in moral philosophy.

It is very important that this mistake is corrected from the start. That mistake, or a failure to correct it in the right way,¹ has resulted in a serious distortion in the philosophy of punishment. Many people reject instrumentalism as an immediate justification of punishment because they regard the implications of consequentialism to be unpalatable. That leads many people to adopt a ‘just deserts’ or retributivist view of punishment. The current popularity of retributivism amongst criminal justice scholars is, I think, at least partly explained in this way.

What we will see is that there are five views that one can hold. Consequentialists and non-consequentialists alike may disagree about whether punishment is intrinsically or instrumentally valuable. So there may be consequentialists who think that punishment is valuable both intrinsically and instrumentally and consequentialists

¹ In *Placing Blame: A Theory of Criminal Law* (Oxford: OUP, 1997) ch.4, for example, Michael Moore recognizes the independence of the intrinsic/instrumental distinction and the consequentialist/non-consequentialist distinction, but almost completely fails to recognize the possibility of a non-consequentialist instrumentalist account of punishment.

who think that it is only valuable instrumentally (obviously enough, there are no consequentialists who think that punishment lacks instrumental value, assuming that punishment has at least *some* good effects). And there may be non-consequentialists who think that punishment is intrinsically valuable, non-consequentialists who think that it is only valuable instrumentally, and non-consequentialists who think that it is valuable both intrinsically and instrumentally.

As I have already said, this book defends the view that punishment can be justified only in virtue of its instrumental value. But it does so in the light of a non-consequentialist view of morality. Demonstrating that this is the right view requires me first to show that punishment is not intrinsically valuable, which will be the task of chapters 3 and 4, and then to justify punishment in the light of a comprehensive non-consequentialist view of morality. That will be the task of the remainder of the book.

Whatever view one takes about the question of the intrinsic value of punishment, probably the most important work to be done in the philosophy of punishment concerns the implications of the more comprehensive philosophy that one adopts for the permissibility of harming others. Those who think that punishment is intrinsically valuable, as well as those who do not, must answer more general questions about the permissibility of harming others. Surprisingly, much of the literature on the philosophy of punishment has not engaged very deeply with these questions.

Here is the structure of this chapter. In section I, I outline a little more clearly the distinction between intrinsic and instrumental value. Section II outlines a retributivist view of punishment and distinguishes it from other views that take the suffering of wrongdoers to have intrinsic value. Section III focuses on the instrumental value and disvalue of punishment: its costs and its benefits. Section IV outlines the distinction between consequentialism and non-consequentialism in more depth. Section V shows that there is no relationship between the immediate justifications of punishment and these more comprehensive philosophical views. The conclusions show the significance that these more comprehensive philosophical views have for the question of punishment regardless of whether one thinks that punishment is intrinsically valuable.

I. Intrinsic and Instrumental Value

What value does punishment have? We may (somewhat crudely) divide the value of things in general into intrinsic and instrumental value. Something is intrinsically valuable if it is valuable in itself. It is intrinsically disvaluable if it has negative value in itself.² Typically at least, we have good reason to pursue things of intrinsic value.

² The distinction is crude, in that strictly speaking intrinsic value may be contrasted not with instrumental value but with extrinsic value, and instrumental value may be contrasted

And we have good reason to avert things that are intrinsically disvaluable. The things that we do that bring about things of intrinsic value, or that avert things that are intrinsically disvaluable, are instrumentally valuable. The things that we do that bring about things of intrinsic disvalue or that avert things of intrinsic value are instrumentally disvaluable. Ultimately, then, the instrumental value and disvalue of things is tied to the intrinsic value and disvalue of things. Some things, of course, have both intrinsic and instrumental value. Going for a walk might be intrinsically valuable in being pleasurable, but also instrumentally valuable in helping one to keep fit.

There can be long chains of instrumentally valuable things that lead eventually to something that is of intrinsic value. For example, suppose that physical pleasure is intrinsically valuable and physical pain is intrinsically disvaluable. Something might be instrumentally valuable because it gives rise to other instrumentally valuable things that, in the end, cause people to feel physical pleasure. Or something might be instrumentally disvaluable because it gives rise to other instrumentally disvaluable things that, in the end, cause people to feel physical pain.

Given that instrumental value is tied to intrinsic value, we have very good reason to develop the right view about what is intrinsically valuable and disvaluable. For only once we have decided that can we determine what is instrumentally valuable and disvaluable. Unfortunately, there is room for a great deal of disagreement about what is intrinsically valuable and it is not very obvious how we are to distinguish what is intrinsically valuable from what is not.

For example, some people might think that it is intrinsically valuable that beautiful flowers survive. The value of the survival of beautiful flowers, they think, does not depend on the pleasure that beautiful flowers might give to people. If that is true, we have reason to keep beautiful flowers alive even if no one will see them. Other people disagree. Some think that the value of beautiful flowers is only instrumental. It is the pleasure that they give to people (and perhaps non-people as well) that makes them valuable. If no one is going to see a beautiful flower there is no reason against destroying it. How do we tell who is right? Not an easy question to answer.

One puzzle at the centre of the philosophy of punishment is about whether the suffering that offenders experience is intrinsically valuable. Some philosophers of punishment claim that it is. It might initially seem puzzling how this could be so. Suffering is generally regarded as a bad thing in itself, even if it is sometimes

not with intrinsic value but with value as a final end. See, further, C Korsgaard 'Two Distinctions in Goodness' in *Creating the Kingdom of Ends* (Cambridge: CUP, 1996) and J Tannenbaum 'Categorizing Goods' in R Shafer-Landau *Oxford Studies in Metaethics* vol.5 (Oxford: OUP, 2008). Furthermore, the intrinsic value of something may be realized only in conjunction with other things. Hence, when I say that something is valuable in itself, I do not mean to imply that it is valuable independently of anything else. These complications need not trouble us here.

necessary to suffer for some further end. Some people might even think that it is definitional of suffering that it is bad for the person concerned. So how could it possibly be intrinsically good for people to suffer?

II. What is Retributivism?

The answer that is often given is that whilst *undeserved* suffering is intrinsically disvaluable, *deserved* suffering is intrinsically valuable. What is intrinsically valuable or disvaluable, it is argued, is relational: it depends on characteristics of the person who gets the thing in question. Good people, or those who do good acts, deserve to be happy. Bad people, or those who do bad acts, deserve to suffer. If a good person, or a person who has done a good act, is happy that is intrinsically valuable and if she suffers that is intrinsically disvaluable. If a bad person, or a person who has done a bad act, is happy that is intrinsically disvaluable and if she suffers that is intrinsically valuable. Whether suffering is good or bad cannot be determined independently of desert. The intrinsic value or disvalue of suffering depends on whether the suffering is deserved.

Another way to put this is to say that the moral valence of happiness and suffering depends on whether the suffering is deserved. Happiness has a positive moral valence if the person suffering is good, but it has a negative moral valence if the person suffering is bad. Suffering has a negative moral valence if the person suffering is good, but it has a positive moral valence if the person suffering is bad. Desert determines the moral valence of suffering. We should pursue things with a positive moral valence and we should avert things with a negative moral valence.

We can specify this idea more precisely by distinguishing the desert basis and the deserved outcome that the desert basis makes valuable. The desert basis of suffering might be morally bad character, or morally wrongful decisions, or morally wrongful actions. For those who think that the moral valence of suffering is determined by desert, suffering is the outcome that is deserved if one or other of these desert bases obtains. In the case of the criminal law, we are concerned with criminal offending (however that is defined) as the desert basis. If that basis is satisfied, some people think, it is valuable that the offender suffers.

How much should the offender suffer? Those who think that the suffering of offenders is intrinsically good typically claim that an offender ought to suffer in proportion to the wrong that he has perpetrated. Suffering that is proportionate to the wrong is good. Suffering that is disproportionate to the wrong is bad. That does not necessarily mean that his suffering ought to be identical to the suffering of the victim, or the intended suffering of the victim. Something can be proportionate to another thing without being identical to that thing (think about the way in which the size of the windows might be disproportionate given the size of the house). Perhaps offenders ought to suffer more than their victims or perhaps less. Those

who think that offenders deserve to suffer will disagree with each other about the degree of suffering that is proportionate for particular wrongs or for wrongs in general.

The idea that wrongdoers deserve to suffer in proportion to the wrong they have committed is a central component of what is commonly called a *retributivist* view of punishment. There is a cluster of views that call themselves retributivist and people who call themselves retributivists believe different things. Retributivists, as I will most often understand the view, believe that criminal wrongdoers deserve to suffer and that the suffering of wrongdoers is intrinsically valuable. They accept that the valence of suffering can be altered by desert.

But the truth of that belief would not be sufficient in itself to justify state punishment. It must also be shown that the person or group doing the punishing is permitted to bring about deserved suffering. Some people might accept that the suffering of offenders is good, but deny that this is a good that the state is permitted to bring about. They might think, for example, that the suffering of wrongdoers is not something that a liberal state can permissibly spend resources on, or they might believe that making wrongdoers suffer would violate their rights.³ These people are not retributivists about state punishment, for they do not think that the fact that offenders deserve to suffer can justify the state in punishing them. Retributivists believe that it is permissible for the state to punish wrongdoers in order to make them suffer, and they believe this because they believe their suffering is deserved.

Hence, I will identify retributivism with two claims. 1) It is intrinsically valuable that offenders suffer in proportion to the gravity of the offence they have committed because that is what they deserve, and 2) it is permissible for the state to ensure that offenders get what they deserve by punishing them.⁴

Retributivists, as I have defined them, might disagree about a further issue. Amongst the things that it is permissible for a person to do, some of those things are morally required. Something is morally required if one has a duty (all things considered) to do that thing. If one has such a duty, one does wrong in not doing that thing. Retributivists typically think not only that the state is *permitted* to punish offenders because they deserve it, but that the state is *required* to punish offenders for that reason. However, retributivism would provide a successful justification of punishment if it could show, more modestly, that punishment is permitted on retributivist grounds.

Retributivists do typically claim other things, in particular that it is *only* permissible to punish offenders in order to give them what they deserve. As we will see this claim is independent of the two claims above. It is to be established by showing that other putative justifications of punishment are false. For this reason, we are best

³ See D Dolinko 'Some Thoughts about Retributivism' (1991) 101 *Ethics* 537 and 'Three Mistakes of Retributivism' (1992) 39 *UCLA Law Review* 1623.

⁴ I will consider a less popular version of retributivism based on rights in chapter 4.

separating out the positive justification of punishment that retributivists offer from the claim that this justification is exclusive of other potential justifications. We might call a person that thinks that it is only permissible to punish offenders for these reasons a *pure* retributivist. A person who believes that the two claims above provide a reason, perhaps a sufficient reason (amongst others), to punish offenders is a retributivist, amongst other things. She holds one kind of ‘mixed’ view of punishment (there are many different mixes and I will make no effort to distinguish between them here).

The standard version of retributivism that I have just outlined can be distinguished from alternative retributive views. These views also hold that the suffering of wrongdoers is intrinsically valuable. For example, some people think that it is intrinsically good that a person is deprived of benefits that she received through a wrongful act. Correcting injustice, we might think, is good in itself. Punishing offenders can bring this good about.⁵ I will discuss this version of retributivism in chapter 4.⁶ We can briefly note a familiar difficulty with it now. Although it might be good that a person who has benefited unjustly from their conduct is deprived of that benefit, doing that does not involve punishing the person. Furthermore, it is difficult to believe that we are *only* permitted to punish people who have received some kind of benefit from their crimes. People who assault other people do not typically benefit from doing that. There is no advantage that the person has typically gained which we now wish to deprive them of.

Other people might think that punishment is justified by its communicative potential. As I noted in the previous chapter, one central function of the criminal justice system is to communicate about wrongdoing. Those who defend punishment in virtue of its communicative potential might disagree about the value of communication. Some may think that it is intrinsically valuable for the community to affirm the rights of victims or to condemn wrongdoers. Others might think that communication has only instrumental value. It is a way of ensuring that victims of crime feel secure and valued so that they can pursue their life projects with confidence.

It is not entirely clear that this communicative view is completely distinct from standard retributivist views, and indeed some who defend a communicative theory

⁵ See H Morris ‘Persons and Punishment’ (1968) 52 *The Monist* 475; J Finnis *Fundamentals of Ethics* (Washington: Georgetown UP, 1983) 127–33.

⁶ Fair play theories of punishment share this kind of problem. For a recent defence, see R Dagger ‘Playing Fair with Punishment’ (1993) 103 *Ethics* (April) 473–88, and ‘Punishment as Fair Play’ (2008) 14 *Res Publica* 259–75. It is very difficult to believe that just punishment is based on unfairly benefiting from a system of rules.

⁷ Morris himself later repudiated his view on these grounds. See ‘A Paternalistic Theory of Punishment’ (1981) 18 *American Philosophical Quarterly* 263. See also A von Hirsch *Censure and Sanctions* (Oxford: OUP, 1993) ch.2; R A Duff *Punishment, Communication, and Community* (Oxford: OUP, 2001) 21–3.

of punishment identify the view as a form of retributivism.⁸ Whether the view should be seen that way depends on whether desert plays a role in justifying the intrinsic value of the communication. If desert does not play this role, the view is distinct from the retributivist claims that I outlined above. If it does, the communicative role should be seen as a variety of retributivism. I will address the communicative view in more detail in chapter 5.

III. The Costs and Benefits of Punishment

Punishment might be justified in the light of its instrumental rather than its intrinsic benefits. Those who think that it is are instrumentalists. Instrumentalists disagree about the range of benefits that can contribute to its justification. Some instrumentalist views are very narrow. Pure paternalism is a very narrow instrumentalist view: pure paternalists claim that the only goods that are relevant to the justification of punishment are goods that are received by the offender herself. We are allowed to punish an offender, pure paternalists claim, only in order to make her better off.⁹ For pure paternalists rehabilitation is central to the immediate justification of punishment.

The offender might benefit in a range of ways by being made to suffer. For one thing, it might be important for her to understand the moral gravity of what she has done. Self-knowledge is plausibly intrinsically valuable, and knowing that one is a wrongdoer might contribute to self-knowledge. But she might also benefit by being provided with guidance for her attitudes and also her future conduct. That might help to repair her relationships with the people that she has wronged, and with the community more generally. Of course, punishment does not have a very strong track record in benefiting offenders in these ways, but paternalists might argue that this is a reason to reform our methods of punishment.

There are some paternalists who are less pure. For them, if punishment is good for the offender that provides one amongst a range of reasons to punish them.¹⁰ Other people deny that punishment, even if our practices are reformed, is likely to do offenders much good.¹¹ Still others think that even if punishing offenders could be good for them that would provide no reason in favour of punishing them. For punishing them for this reason would pursue their good only by failing to respect their autonomy.¹² On this view, if any other value that punishment has is

⁸ See, for example, A von Hirsch *Censure and Sanctions* ch.2.

⁹ See J Hampton, 'The Moral Education Theory of Punishment' (1984) 13 *Philosophy of Public Affairs* 208.

¹⁰ See, for example, H Morris 'A Paternalistic Theory of Punishment' (1981) 18 *American Philosophical Quarterly* 263.

¹¹ See R Schafer-Landau 'Can Punishment Morally Educate?' (1991) 10 *Law and Philosophy* 189.

¹² See, further, chapter 15.

insufficient to warrant punishing the offender, any good that it does to the offender cannot render punishment permissible. This might be thought true for the reason that it is not permissible to harm a person for their own good.

Another approach connects paternalist ambitions to a communicative view of punishment, a view that sees communicating with the offender and the community by imposing harsh treatment on the offender as intrinsically valuable if it is done for the sake of the offender.¹³ On this view, we owe it to the offender, the victim, and the political community to ensure that the offender comes to recognize that what she has done is wrong and to reconcile herself with the victim in particular and her fellow citizens more generally. Our obligations to the offender play an important role in justifying punishment on this view.

Over and above any good that it might do to the offender herself, the most obvious good that punishment might do is to reduce crime. Crime reduction can come about through punishment in a number of ways. It might come about because the offender, through being punished, is incapacitated. If the offender is in prison he cannot commit further crimes on the outside for the period of his imprisonment. But, of course, the offender can commit crimes on the inside. Those people who argue that 'prison works' often (disgracefully) neglect this fact about crime. They may think that crime on the inside is not morally significant, perhaps because prisoners deserve to be victims of crime, but more likely they do not pay any attention to the problem of crime in prison at all because it is unlikely to show up on crime surveys and because reducing the rate of crime in prison is hardly a vote-winner.

Incapacitation alone, though, is not punishment. We can incapacitate a person without having any intention that the person suffer, as we sometimes do with dangerous people who suffer from a serious mental disorder. We could incapacitate offenders and provide them with compensation by ensuring that their cells are large, beautifully furnished, and full of rewarding things to do. Were we to do that we would not be punishing offenders, we would *merely* be incapacitating them. Incapacitation, for this reason, is probably not best thought of as part of the justification of punishment, though particular modes of punishment may be justified on the grounds that they also prevent crime through incapacitation.

Crime reduction might also come about because the offender himself will be deterred from committing further crimes by being punished. An offender who is punished may think again about committing further crimes. This is often called *special deterrence*. It can be distinguished from *general deterrence*. General deterrence involves the punishment of offenders to deter other people from committing crimes. People who would otherwise be tempted to commit crimes may respond to the fact that there is a threat of punishment, particularly if they see that this threat is regularly carried out against offenders.

¹³ This is a way of reading R A Duff *Punishment, Communication, and Community*.

Some people think that it is permissible to punish offenders for reasons of special deterrence, but that it is wrong to punish offenders for reasons of general deterrence.¹⁴ They might think this because the offender is, for the most part, not responsible for the threats that other potential offenders pose to other potential victims. They may think that it is wrong to use the offender as a means to avert these threats. Others think that some degree or kind of general deterrence can contribute to the justification of punishment. They might think that it can contribute only if, and to the extent that, the offender makes her victim, or others, more vulnerable to future attacks and that she can be harmed to avert those attacks,¹⁵ or they might think that it can play a role beyond this in the justification of punishment.

But these are not the only potential benefits of punishment. Punishment is also an industry. The building, running, governing, and study of punishment systems all provide employment which may give people job satisfaction and contribute to the flourishing of the economy. Furthermore, punishment may contribute to art. Some very good books have been written in and about prisons and there are some great prison movies. Do these benefits also contribute to the justification of punishment?¹⁶

For instrumentalists, it is very important to identify what valuable things can contribute to the justification of punishment. For punishment is also very costly. Even leaving aside the suffering that the offender might experience, punishment prevents offenders doing valuable things that they might have done had they not been punished, by depriving them of the resources or liberty that they would otherwise have had. It destroys and hinders the development of relationships between the offender and his family and friends, many of whom will feel ashamed and distressed at the punishment of the offender. It is very expensive, and it takes offenders out of the workforce. And just as it may deter crime, so it may stimulate crime by helping offenders to learn from each other and by destroying their career opportunities so that crime becomes relatively attractive to them. Furthermore, it may tend to entrench racism and other forms of discrimination. And any system of punishment will result in some people who are innocent being punished.

The way in which the costs and benefits figure in the justification of punishment is a question as much for retributivists as for instrumentalists. It is disappointing that

¹⁴ See, for example, W Quinn 'The Right to Threaten and the Right to Punish' in *Morality and Action* (Cambridge: CUP, 1993).

¹⁵ See D M Farrell 'The Justification of Deterrent Violence' (1990) 100 *Ethics* 301.

¹⁶ For a very permissive view of the range of benefits that might contribute to the justification of punishment, see J Gardner 'Crime: In Proportion and In Perspective' in *Offences and Defences: Selected Essays in the Philosophy of Criminal Law* (Oxford: OUP, 2007). In the context of just war theory, some people think that benefits of this kind cannot help to make going to war proportionate. See T Hurka 'Proportionality and Necessity' in L May *War: Essays in Political Philosophy* (Cambridge: CUP, 2008). I will discuss this issue in detail in chapter 15.

some of the most prominent retributivists have done so little to develop a plausible account of the way in which instrumental value and disvalue figure in their views about punishment. For example, Michael Moore recognizes that punishment may have good effects. But on his view these good effects play no role whatsoever in justifying punishment. They are simply 'a happy surplus that punishment produces and form no part of what makes punishment just'.¹⁷

But it is very difficult to believe that the good consequences of punishment play *no* role in the justification of punishment. As there are many reasons *against* punishing people the benefits of punishment must at least be important in offsetting these bad consequences of punishment.¹⁸ Were that not the case, retributivists would have to show either that the instrumental reasons against punishment do not figure in the justification of punishment either, or that punishment is justified only if the good of giving wrongdoers what they deserve outweighs these bad consequences of punishment. Neither of these views is at all plausible.

There is, then, a lot of work for retributivists and instrumentalists alike to do in exploring the relationship between the costs and benefits of punishment. For a start, we might think that it is worse to cause harm than to fail to prevent harm. If that is true, the extent to which the criminal justice system is effective in reducing the crime rate must be of significantly greater value than the harm that the system causes for it to be justified. Furthermore, we might wonder whether all costs and all benefits are relevant in determining whether our criminal justice institutions are just. We might think that it matters who must bear these costs, the way in which they are distributed, what kinds of cost they are, and the way they are brought about.

Whilst scholars working on distributive justice have explored many related issues with depth and sophistication, philosophers of punishment have made very little progress on them. Philosophers of punishment have been so caught up with defending or attacking retributivism that these very morally significant and difficult issues have been left unaddressed. The truth or falsity of retributivism is undoubtedly important. But it is a modest component of a full theory of punishment.

IV. Consequentialism and Non-Consequentialism

How we should go about working out the way in which the costs and benefits of punishment can figure in its justification depends on the comprehensive moral view that one holds. In moral philosophy there is an important distinction between

¹⁷ *Placing Blame* 153.

¹⁸ See, especially, D N Husak 'Retribution in Criminal Theory' (2000) 37 *San Diego Law Review* 959, 974 and also M Berman 'Punishment and Justification' (2008) 118 *Ethics* 258.

consequentialist and non-consequentialist views. Consequentialists believe that the only thing that matters in determining whether an action, rule, or practice is right or wrong is the consequences of that action, rule, or practice. Non-consequentialism is the view that the consequences of an action, rule, or practice are not all that matter in determining what is right or wrong.

As should already be clear, there are many different consequentialist and non-consequentialist views that one might hold. Consequentialists typically believe that what matters in determining what is right and wrong is the extent to which an action brings about things that are intrinsically good. They disagree, though, about what is intrinsically valuable. The simplest view, but one that is still prominent in writing in the philosophy of punishment, is utilitarianism. Utilitarians believe that a very narrow range of consequences is valuable. Typically they focus on happiness and unhappiness or pleasure and pain, but sometimes also on preference satisfaction. Other consequentialists think that these are not the most important consequences. They might think that worthwhile actions autonomously pursued are the good consequences that we should pursue,¹⁹ or they might think that freedom from arbitrary control of others is a good consequence to aim at.²⁰

Consequentialists might also disagree about whether the way in which valuable things are distributed matters morally. Classical utilitarians, building on the views of Jeremy Bentham, took the view that what matters is maximizing happiness or pleasure and minimizing unhappiness or pain. But consequentialists need not be committed to this maximizing view. Some might believe that a more equal distribution of valuable things amongst people is better than a less equal distribution. The way in which valuable things are distributed, consequentialists might believe, contributes to the overall magnitude of value attained.²¹ But whatever they believe about the quality of the consequences or their distribution, consequentialists believe that the consequences are all that matter in deciding what is right and what is wrong.

Many non-consequentialists think that there are restrictions on the pursuit of good consequences and permissions not to pursue good consequences. But they disagree about the source of these restrictions and permissions as well as their content. Building on the pioneering work of Immanuel Kant, non-consequentialists typically think that the moral status that human beings have is important in determining the principles that ought to guide our practical reasoning. The respect that we owe to each other in virtue of our capacity to critically evaluate and determine our own ends constrains what we may do to each other and what we owe to each other.

¹⁹ See, for example, J Raz *The Morality of Freedom* (Oxford: OUP, 1986) ch.12. It is not completely clear to me whether Raz is best characterized as a consequentialist.

²⁰ See P Pettit *Republicanism: A Theory of Freedom and Government* (Oxford: OUP, 1997).

²¹ This is made clear in D Parfit 'Equality and Priority' in M Clayton and A Williams *The Ideal of Equality* (London: Palgrave, 2002).

This abstract idea and its implications, both at a fundamental theoretical level and also in terms of general principle, are also highly controversial. Non-consequentialists typically endorse at least two principles that form part of our ordinary morality which consequentialists reject. Part of the attraction of non-consequentialism is that it may be able to explain these features of ordinary morality that it is difficult for us to give up. First, they endorse the *means principle*. They believe that it is wrong to harm a person as a means even if doing so produces a greater good. This principle is intuitively attractive. For example, many people believe that it would be wrong to kill me even if the result of doing so would be that you could distribute my healthy organs to five other people who need them in order to survive. That is so even though killing me to save the five has a net saving of four lives.

Secondly, they endorse the *doctrine of acts and omissions*. They believe that it is permissible not to aid a person who will be harmed even if the harm that the person would have to bear in doing so is less serious than the harm that she would avert. For example, suppose that a stranger is caught underneath a machine and is being crushed to death. I realize that I can save him. But if I do this my arm will be severed by the machine. Many people believe that I am permitted not to save him even though the harm that he will otherwise suffer (death) is greater than the harm that I would suffer (loss of an arm) were I to save him. Saving the stranger would be morally good, but it is not morally required. To save him would be beyond the call of duty.

We can summarize the distinction between consequentialists and non-consequentialists that we have outlined so far in the following way: many non-consequentialists, in contrast with consequentialists, believe that there are restrictions on pursuing a greater good, such as the *means principle*, as well as permissions not to pursue a greatest good, such as the *doctrine of acts and omissions*.

Consequentialists and non-consequentialists may also come apart about the kinds of consequences that can figure in certain kinds of decisions. Consequentialists will typically think that all of the intrinsically good consequences of a decision can militate in favour of that decision, in principle at least. If we exclude some consequences from our deliberations we do that for pragmatic reasons. People may be more likely to achieve good consequences if they do not always focus on all of the consequences of their actions in their decisions.²² We make better decisions by limiting our focus to the things that are the most important. In principle, though, all of the consequences of our actions ought to be taken into consideration in determining whether those actions are right or wrong.

²² See, especially, P Railton 'Alienation, Consequentialism, and the Demands of Morality' in *Facts, Values, and Norms* (Cambridge: CUP, 2003). For an exploration of a somewhat similar view in the context of the criminal law, see J Gardner 'Complicity and Causality' in *Offences and Defences*. Gardner describes his view as 'agent-neutralist' and he contrasts it with consequentialism. But as we shall see in chapter 6 it is better thought of as a form of consequentialism that regards wrongful actions as intrinsically disvaluable.

In contrast, some non-consequentialists think that there are some costs and benefits that cannot contribute to some decisions in principle. They are excluded not because we tend to produce worse consequences by thinking about them, but rather because it is wrong in principle to be guided by them in making certain decisions. For example, if I am deciding whether to save one stranger or another, both of whom are in mortal danger, I may not be permitted to take into consideration some of the good consequences that might be brought about if I decide one way rather than another. Suppose that other things are equal, but if I saved one person I would at the same time save a pretty flower that he is carrying. Some non-consequentialists argue that there are reasons in principle why the saving of the flower is irrelevant in determining whom to save in a case like this. They argue that I am not permitted to take the saving of the flower, valuable though that is, into consideration in deciding whom to save. I ought rather to toss a coin.²³ Consequentialists will endorse this conclusion, if they do, only for pragmatic reasons rather than reasons of principle. In principle, they will suggest, we should save the person with the flower.

As I noted in the previous section, even if we agree that the instrumental benefits of punishment can contribute to its justification, we might disagree about which instrumental benefits can figure in our decision to set up institutions of punishment and our decisions about whom to punish and to what degree. For example, we might wonder whether the satisfaction that prison officers feel in doing their job can contribute to the justification of punishment, even if we think that it is valuable that they feel that way. We might wonder whether the fact that punishing a particular offender would provide an opportunity for another person to do the offender's job, a person who might do it better than the offender would, is a reason in favour of punishing him. Instrumentalists face a deep and difficult task in determining which consequences are relevant to the justification of punishment and why, both in general and in particular cases.

V. The Immediate and the Deep Justifications of Punishment

In the philosophy of punishment instrumentalist views are often associated with consequentialist philosophy and retributivist views are often associated with non-consequentialist philosophy. These relationships have a historical pedigree. The forefather of consequentialist thinking, Jeremy Bentham, provided an instrumentalist justification of punishment. And indeed that is the only view that is consistent with his particular brand of consequentialism. Immanuel Kant, the forefather of non-

²³ See, especially, F M Kamm *Morality, Mortality I: Death and Whom to Save from It* (Oxford: OUP, 1993) ch.8.

consequentialist thinking, provided a retributivist justification of punishment. However, despite this historical pedigree, both the relationship between consequentialism and instrumentalism and the relationship between non-consequentialism and retributivism are philosophically weak.²⁴

Much of the philosophy of punishment is confused by the assumption that there is good reason to think that these relationships are significant. A standard objection to 'mixed' theories of punishment is that they pick and mix the parts of different philosophical views that fit their liberal intuitions.²⁵ It is supposed that there is something amiss if one is an instrumentalist and at the same time one thinks that there are philosophical limits on the pursuit of instrumental goals. It is true that some mixed theories of punishment, notably that defended by H L A Hart,²⁶ do seem philosophically incoherent in this way. But as a general claim about views that are both instrumentalist and non-consequentialist, this thought is thoroughly misguided. Given the prevalence of this objection, it is worth spending some time clearing up the confusion.

First, let us focus on retributivism. Retributivism might be defended either by consequentialists or by non-consequentialists.²⁷ Retributivism holds that the suffering of wrongdoers is intrinsically valuable. But both consequentialists and non-consequentialists alike must agree that there are things that are intrinsically valuable, and may agree about what is intrinsically valuable. The debate between consequentialists and non-consequentialists is not a debate about what is intrinsically good. They disagree not about which things are intrinsically good, but about whether the consequences, the bringing about of things that are intrinsically good, are all that matter in determining what is right or wrong. Furthermore, it may well be that retributivism is more appealing to consequentialists than to non-consequentialists. This is because, for consequentialists, there is no special difficulty in the idea that the state should promote outcomes which are intrinsically good. For them there are no special restrictions on the pursuit of good consequences. For non-consequentialists, in contrast, it does not follow from the idea that the suffering of wrongdoers is intrinsically good that the state may pursue that good. If pursuing the good amounts to coercing citizens, it may be that the state cannot pursue the good. To do so would be to coerce citizens as a means to the good.

This point, and its significance, is often missed. Many retributivists think that it is a virtue of their view that it demonstrates why it is wrong to punish the innocent. But retributivism, even were it the right view, would not establish this. It holds only

²⁴ See also, as a prelude to a discussion of punishment very different from mine, J Tasioulas 'Punishment and Repentance' (2006) 81 *Philosophy* 279.

²⁵ Some people who defend mixed theories accept this difficulty, and then endorse it as a virtue. See, for example, D Wood 'Retribution, Crime Reduction and Justification' (2002) 22 *Oxford Journal of Legal Studies* 301.

²⁶ *Punishment and Responsibility* 2nd edn (Oxford: OUP, 2008).

²⁷ See also M Moore *Placing Blame* 155–9.

that the intrinsic goodness of the suffering of wrongdoers is a good reason to punish them. Whether there is sufficient reason to punish people who *don't* deserve to suffer is a question that is completely independent of the truth of retributivism. If there are restrictions on the permissibility of punishing the innocent for the sake of a greater good, these restrictions have nothing at all to do with desert. It is true that if we punish the innocent we would punish people who don't deserve to be punished. This fact must count against punishing them in the eyes of retributivists—their suffering is intrinsically bad. But retributivism has no implications for whether it is permissible bring about something that is intrinsically bad for the sake of the greater good. Given that crime *also* makes innocent people suffer undeservingly, punishing the innocent, were it to have deterrent effects, might maximize deserved outcomes. In other words, the truth of retributivism would not establish that retributivism provides the *only* justification of punishment. The idea of desert might be used to justify general deterrence as an immediate aim of punishment. If there are restrictions on punishing the innocent to prevent victims of crime suffering undeserved harms, these restrictions have nothing to do with the idea of desert itself.

The same thing holds true for the question of proportionality. Many retributivists think that it is a virtue of their view that it shows that it is wrong to punish offenders in a way that is disproportionate to the wrong that they have committed, even if that would achieve a greater good. But retributivism would not establish the truth of this either. It holds only that there is a good reason to punish offenders in proportion to the wrong that the person has committed. Demonstrating that to be true would not imply that it is wrong to punish offenders more than they deserve if this would achieve a greater good. The truth of retributivism would not show that we could not punish people disproportionately for *other* reasons—most obviously to ensure that victims of crime do not suffer disproportionately as a result of being offended against.

This is also a problem for more 'mixed' views. John Tasioulas, for example, thinks that retributivism sets upper and lower limits for punishment, but that instrumental considerations might be important in determining how much to punish within the permissible range.²⁸ But he offers no argument for the view that we cannot punish a person more than she deserves for instrumental reasons grounded in desert. The idea that retributivism can provide a constraint on how much it is permissible to punish is wrongheaded. A positive reason to do something cannot provide a constraint on doing that thing.

Of course many retributivists do claim that it is *only* permissible to punish a person because she deserves punishment, and that it is *only* permissible to punish a person in proportion to the crime that she has committed. But these are mere assertions in need of a defence. At this point it is common for retributivists simply to utter a few non-consequentialist phrases about not treating people as a means or treating people

²⁸ 'Punishment and Repentance'.

as people rather than as animals. But they have done little to investigate the extent to which these ideas actually provide constraints on the practices of punishment. Perhaps they think that it is never permissible to harm a person intentionally without that person's consent for a greater good. But that view is wholly implausible whether or not one is a consequentialist, as we can see from the fact that it is permissible to defend oneself or another person against a culpable attacker.

Defending the two retributivist claims that I outlined in section II, then, does nothing at all to advance the cause of those who wish to justify constraints on punishing the innocent and punishing disproportionately. The truth of the retributivism would give us a positive reason *for* punishing wrongdoers proportionately. And that reason would apply only to wrongdoers. But it would give us no reason *against* punishing the innocent or against punishing wrongdoers disproportionately. This just follows from the very simple truth that the existence of a reason to do something does nothing to exclude the possibility that the thing (and more) could not adequately be justified on other grounds.

Just as a retributivist might be either a consequentialist or a non-consequentialist, so an instrumentalist may be either a consequentialist or a non-consequentialist. Consequentialists and non-consequentialists agree that the consequences of our actions matter in determining whether those actions are right or wrong. They disagree about whether they are the *only* things that matter. Non-consequentialists typically believe that there are restrictions on punishing the innocent for the sake of the greater good. They endorse this view whether or not they are retributivists. Consequentialists typically believe that there are no such restrictions. They endorse that view whether or not they are retributivists.

It is common to reject instrumentalist views of punishment on the grounds that instrumentalists regard it as permissible to punish the innocent for the sake of the greater good. The classic utilitarian accounts of punishment, such as John Rawls's early work,²⁹ have provided quite forceful arguments that the punishment of the innocent wouldn't be very effective in promoting the greater good. But as will now be clear, even if those utilitarians are wrong, and punishing the innocent could achieve the greater good, an instrumentalist may or may not believe that this is permissible. That depends on whether the instrumentalist is also a consequentialist. A person who believes that punishment is justified by its good effects might also believe that there are restrictions on the pursuit of these good effects. If non-consequentialist philosophy can provide a justification for the constraints on punishment that many of us find attractive, that justification will be available to retributivists and instrumentalists alike. The difference between the retributivist and the instrumentalist is a difference about the positive reasons for punishment, not about the constraints on our practices.

²⁹ 'Two Concepts of Rules' in *Collected Papers* (S Freeman (ed)) (Cambridge, Mass.: Harvard UP, 1999).

It is surprising that so many people assume that if we reject consequentialism we must also reject instrumentalism. Even many instrumentalists appear to assume that instrumentalism and consequentialism must go together.³⁰ This is surprising because very significant progress has been made in outlining a range of different non-consequentialist views of morality since the pioneering work of Rawls.³¹ The idea of moral desert has almost no role at all to play in this scholarship, and some of the most prominent non-consequentialist philosophers either reject the idea of moral desert altogether or fail to consider it in any depth.³² One central theme of that literature has been the justification of constraints on the pursuit of good consequences.

Even some Kantians reject retributivism, on the grounds that the intuitionist basis of retributivism is at odds with a range of other principles and moral ideas that Kant endorses.³³ Some even think (less plausibly) that Kant himself was not a retributivist.³⁴ Given that, one might expect the conjunction of non-consequentialism and instrumentalism to be more popular amongst philosophers of punishment.

What we can see from this sketch of the terrain in the philosophy of punishment is that we have two different debates that need to be resolved in order to justify a set of punishment practices. One debate concerns the question of whether the suffering of wrongdoers is intrinsically valuable or disvaluable, and the extent to which punishment can be justified on this basis. That is a debate about what positive reasons there are to punish. Retributivists think that the intrinsic value of the suffering of wrongdoers is a positive reason to punish them. The next two chapters investigate whether retributivists are right to think this. Instrumentalists, in contrast, typically think that the suffering of wrongdoers is a reason against punishing them. Instrumentalists think that crime reduction is a good reason to punish offenders. There is nothing in retributivism that addresses this issue one way or another, and retributivists may have different views about it. Some retributivists might think that punishment may be justified in different ways.

³⁰ That is true even for those such as David Dolinko, who understand that retributivism and non-consequentialism need not go together. See 'Retributivism, Consequentialism, and the Intrinsic Goodness of Punishment' (1997) 16 *Law and Philosophy* 507.

³¹ In *A Theory of Justice* revised edn (Oxford: OUP, 1999) Rawls famously dismisses the role of desert in delivering distributive justice. Whether he accepts the importance of moral desert in punishment is a more difficult question to answer. See S Scheffler 'Justice and Desert in Liberal Theory' in *Boundaries and Allegiances: Problems of Responsibility and Justice in Liberal Thought* (Oxford: OUP, 2000).

³² See, for example, T M Scanlon *What We Owe to Each Other* (Cambridge, Mass.: Harvard UP, 1998) 274–7. More recently Scanlon has endorsed a much more limited account of moral desert, though one which could not play a role in the justification of punishment. See *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, Mass.: Harvard UP, 2008) 188–90.

³³ See A Wood *Kantian Ethics* (Cambridge: CUP, 2008) ch.12.

³⁴ See S Byrd 'Kant's Theory of Punishment: Deterrence in its Threat, Retribution in its Execution' (1989) 8 *Law and Philosophy* 151; T Hill *Human Welfare and Moral Worth* (Oxford: OUP, 2002) ch.10. As Wood (ibid.) demonstrates, evidence of Kant's retributivism is amply supported by the texts.

Another debate concerns the question whether consequentialism or non-consequentialism is the better moral view, and the implications of the answer to this question for restrictions on punishment. Some consequentialists, but not all, think that the suffering of the offender is a reason against punishing him. For whilst some consequentialists will think that a person's suffering always has a negative valence, other consequentialists disagree. It is consistent with consequentialism to believe that deserved suffering is intrinsically good.³⁵ All consequentialists, of course, agree that crime reduction is a good reason to impose punishment on offenders. For surely it is a bad thing, a thing to be averted if we can, that other people are victims of crime.

Non-consequentialists, in contrast, may disagree whether crime reduction is a reason to punish offenders. Some non-consequentialists might think that it is wrong to harm offenders as a means to reduce crime. They may think that this violates the *means principle*, and that harming offenders for this reason fails to treat people with the respect that they are due as morally autonomous agents. But other non-consequentialists will think that offenders can be punished, at least for some instrumental aims and within certain constraints.

This last point suggests that all philosophers of punishment ought to investigate non-consequentialist principles and their implications. That investigation is important whether or not one is a retributivist. Retributivists and instrumentalists are in more or less the same boat when it comes to defending the liberal idea that we must only punish the guilty and we must only punish people proportionately. Both have to explain why it is that it is wrong to harm the innocent, or harm the guilty disproportionately, as a means to a greater good.

Conclusions

We have seen that there are two main questions that need to be addressed in the justification of punishment. One question is whether the suffering of offenders is intrinsically good. The other question, the deeper and more difficult question, is whether we should endorse a consequentialist or a non-consequentialist view of the permissibility of harming. This second question is essential to the justification of punishment, regardless of the view that one takes about the first question.

With respect to the first question, I will argue that the value of punishment is completely instrumental. There is nothing good in itself about making offenders suffer. Rejecting retributivism will occupy us in the next two chapters. Parts II, III,

³⁵ For example, Shelly Kagan is an advocate of both consequentialism and the importance of desert. For his consequentialism, see *The Limits of Morality* (Oxford: OUP, 1989). His work on desert is being developed. But see 'Equality and Desert' in L P Pojman and O McLeod *What Do We Deserve? A Reader on Justice and Desert* (Oxford: OUP, 1999) and 'Comparative Desert' in S Olsaretti *Desert and Justice* (Oxford: OUP, 2003).

and IV of the book are focused on the more difficult, and I think more important, second question. This is not a question that has typically received very much attention in the philosophy of punishment, which has been all too consumed with the first question. Fortunately, there is a great deal of work on which to draw in helping us to make progress in answering it. The theory of punishment that I will defend, then, is both instrumentalist and non-consequentialist. It is instrumentalist in that we are permitted to punish offenders only because of the good consequences that punishment will have. And those good consequences are primarily good for other people, including the victim, who would otherwise be at greater risk of being treated wrongfully, either by the offender himself or by others. Punishment is justified only insofar as it is effective in reducing crime.

Furthermore it is justified only insofar as it is more effective in reducing crime than other less costly and harmful methods that we might use. Given that punishment harms both offenders and non-offenders, and given that it is very costly, there is every reason to seek alternatives to punishment. The theory of punishment that I offer may, in that way, be a stepping stone to abolition.³⁶ Perhaps there are no permissible alternatives to punishment that would be satisfactory in terms of crime reduction or cost. Whether that is true raises empirical questions that I cannot answer.

The duty view of punishment that I offer is non-consequentialist, in that it is developed in the light of moral principles that provide restrictions on pursuing good consequences. I defend the two central non-consequentialist principles, the *means principle* and the *doctrine of acts and omissions*. I defend these on the basis of a deeper idea that lies at the heart of non-consequentialism as a moral philosophy: the importance of the status that autonomous agents have. We will see that these principles are important, but that they have their limits.

Instrumentalist defences of punishment are committed to harming some people as a means to the good of others. The question is whether this is consistent with the best understanding of the *means principle*. I will show that the *means principle* is not an absolute principle. We can be morally liable to be treated as a means to pursue a good, and wrongfully harming others is a central way in which such liability is generated. This helps to explain why the innocent may not be punished—they are not liable to be harmed as a means, for they have no special duty to pursue the good. It also explains why punishment must be proportionate—there are limits to what a person is liable to suffer as a result of their wrongdoing. Familiar restrictions on punishment, then, are best understood not as a consequence of retributivism, but rather as a consequence of the requirement that it is only the liable that are harmed as a means to the good. But these are only hints at an argument to come.

³⁶ For a recent defence of abolitionism, see D Boonin *The Problem of Punishment* (Cambridge: CUP, 2008).