Legal Punishment

First published Tue Jan 2, 2001; substantive revision Tue Jul 18, 2017

The question of whether, and how, legal punishment can be justified has long been a central concern of legal, moral, and political philosophy: what could justify a state in using the apparatus of the law to inflict intentionally burdensome treatment on its citizens? Radically different answers to this question are offered by consequentialist and by retributivist theorists — and by those who seek to incorporate consequentialist and retributivist considerations in ‘mixed’ theories of punishment. Meanwhile, abolitionist theorists argue that we should aim to replace legal punishment rather than to justify it. Among the significant developments in recent work on punishment theory are the characterisation of punishment as a communicative enterprise, greater recognition that punishment’s justification depends on the justification of the criminal law more generally, and growing interest in the normative challenges raised by punishment in the international context.

1. Legal Punishment and its Justification
2. Punishment, Crime, and the State
3. Consequentialist Accounts
4. Retributivist Accounts
5. Punishment as Communication
6. Mixed Accounts
7. ‘Restorative Justice’ and Restitution
8. Further Issues
Bibliography
Academic Tools
Other Internet Resources
Related Entries

1. Legal Punishment and its Justification

The central question asked by philosophers of punishment is: What can justify punishment? More precisely, since they do not usually talk much about punishment in such contexts as the family or the workplace (but see Zaibert 2006; Bennett 2008: Part II), their question is: What can justify formal, legal punishment imposed by the state on those convicted of committing criminal offences? We will also focus on legal punishment here: not because the other species of punishment do not raise important normative questions (they do), nor because such questions can be answered in terms of an initial justification of legal punishment as being the paradigm case (since it’s not clear that they can be), but because legal punishment, apart from being more dramatically coercive and burdensome than other species of punishment usually are, raises distinctive issues about the role of the state and its relationship to its citizens, and about the role of the criminal law. Future references to ‘punishment’ should therefore be read, unless otherwise specified, as references to legal or criminal punishment.

What then are we to justify in justifying punishment? The search for a precise definition of punishment that exercised some philosophers (for discussion and references, see Scheid 1980; Boonin 2008: 3–28; Zimmerman 2011: ch. 1) is likely to prove futile: but we can say that legal punishment involves the imposition of something that is intended to be both burdensome and reprobative, on a supposed offender for a supposed crime, by a person or body who claims the authority to do so. Two points are worth particular notice here.

First, punishment involves material impositions or exactions that are in themselves typically unwelcome: they deprive people of things that they value (liberty, money, time); they require people to do things that they would not normally want to do or do voluntarily (to spend time on unpaid community labour, to report to a probation officer regularly, to
How far it matters, in this context, to make explicit a political theory of the state depends on how far different plausible

law at all?) and of the proper powers and functions of the state (by what authority or right does the state make and

presuppose, and should perhaps make explicit, a normative account of the criminal law (why should we have a criminal

the law and to impose punishments. A normative account of legal punishment and its justification must thus at least

2. Punishment, Crime, and the State

 Legal punishment presupposes crime as that for which punishment is imposed, and a criminal law as that which defines

crimes as crimes; a system of criminal law presupposes a state, which has the political authority to make and enforce

the law and to impose punishments. A normative account of legal punishment and its justification must thus at least

presuppose, and should perhaps make explicit, a normative account of the criminal law (why should we have a criminal

law at all?) and of the proper powers and functions of the state (by what authority or right does the state make and

declare law, and impose punishments on those who break it?).

How far it matters, in this context, to make explicit a political theory of the state depends on how far different plausible
political theories will generate very different accounts of how punishment can be justified and should be used. We cannot pursue this question here (for two sharply contrasting views on it, see Phillips 1986, Davis 1989; for more recent contributions showing the importance of political theory, see Pettit 1997; Matravers 2000; Dolovich 2004; Garvey 2004; Dagger 2007; Brettschneider 2007; Sigler 2011; Markel 2012; Chiao 2016; Flanders 2017), save to note one central point. For any political theory (most obviously any version of liberalism or republicanism) that takes seriously the idea of citizenship as full membership of the polity, the problem of punishment takes a particularly acute form, since we have now to ask how punishment can be consistent with citizenship (how citizens can legitimately punish each other); if we are not to say that those who commit crimes thereby forfeit their status as citizens (see s. 6 below), we must — if we are to justify punishment at all — show how the imposition of punishment can be consistent with, or even expressive of, the respect that citizens owe to each other. (Punishment is also, of course, imposed on non-citizens who commit crimes within a state’s territory: on the primacy of citizenship in understanding criminal law and its authority, and on the status of non-citizens, see Duff 2013.)

Before we tackle such theories of punishment, however, we should look briefly at the concept of crime, since that is one focus of the abolitionist critique of punishment.

On a simple positivist view of law, crimes are kinds of conduct that are prohibited, on pain of threatened sanctions, by the law; and for positivists such as Bentham, who combine positivism with a normative consequentialism, the questions of whether we should maintain a criminal law at all, and of what kinds of conduct should be criminalised, are to be answered by trying to determine whether and when this method of controlling human conduct is likely to produce a net increase in good. Such a perspective seems inadequate, however: inadequate both to the claims of the criminal law, which presents its demands as something other or more than those of a gunman writ large — as something other or more than ‘Behave thus, or else!’ — and to the normative issues at stake when we ask what kinds of conduct should be criminalised. For the criminal law portrays crime not merely as conduct which has been prohibited, but as a species of wrongdoing: whether our inquiry is analytical (into the concept of crime) or normative (as to what kinds of conduct, if any, should be criminal), we must therefore focus on that notion of wrongdoing.

Crimes are, at least, socially proscribed wrongs — kinds of conduct that are condemned as wrong by some purportedly authoritative social norm. That is to say that they are wrongs which are not merely ‘private’ affairs, which properly concern only those directly involved in them: the community as a whole — in this case, the political community speaking through the law — claims the right to declare them to be wrongs. But crimes are ‘public’ wrongs in a sense that goes beyond this. Tort law, for instance, deals in part with wrongs that are non-private in that they are legally and socially declared as wrongs — with the wrong constituted by libel, for instance. But they are still treated as ‘private’ wrongs in the sense that it is up to the person who was wronged to seek legal redress. She must decide to bring, or not to bring, a civil case against the person who wronged her; and although she can appeal to the law to protect her rights, the case is still between her and the defendant. By contrast, a criminal case is between the whole political community — the state or the people — and the defendant: the wrong is ‘public’ in the sense that it is one for which the wrongdoer must answer not just to the individual victim, but to the whole polity through its criminal courts.

It is notoriously difficult to give a clear and plausible account of the distinction between civil and criminal law, between ‘private’ and ‘public’ legal wrongs, whether our interest is in the analytical question of what the distinction amounts to, or in the normative question of which kinds of wrong should fall into which category (see Murphy and Coleman 1984, ch. 3; a symposium in Boston University Law Review vol. 76 (1996): 1–373; Lamond 2007). It might be tempting to say that crimes are ‘public’ wrongs in the sense that they injure the whole community: they threaten social order, for instance, or cause ‘social volatility’ (Becker 1974); or they involve taking unfair advantage over those who obey the law (Murphy 1973, Dagger 1993); or they undermine the trust on which social life depends (Dimock 1997). But such accounts distract our attention from the wrongs done to the individual victims that most crimes have, when it is those wrongs that should be our central concern: we should condemn the rapist or murderer, we should see the wrong he has done as our concern, because of what he has done to his victim. Another suggestion is that ‘public’ wrongs are those that flout the community’s essential or most basic values, in which all members of the community should see themselves as sharing: the wrong is done to ‘us’, not merely to its individual victim, in the sense that we identify ourselves with the victim as a fellow citizen (see Marshall and Duff 1998; Duff 2007, ch. 6; and see further section 6 of the entry on theories of criminal law).

Some abolitionists, however, argue that we should seek to eliminate the concept of crime from our social vocabulary: we should talk and think not of ‘crimes’, but of ‘conflicts’ or ‘troubles’ (Christie 1977; Hulsmann 1986). One motivation for this might be the thought that ‘crime’ entails punishment as the appropriate response: but that is not so, since we could imagine a system of criminal law without punishment. To define something as a ‘crime’ does indeed imply that some kind of public response is appropriate, since it is to define it as a kind of wrong that properly concerns the whole community; and it implies that that response should be a condemnatory one, since to identify wrongs as wrongs is to mark them out as apt for condemnation: but that public, condemnatory response could consist in nothing more than, for instance, some version of a criminal trial which calls the alleged wrongdoer to answer for her alleged wrongdoing, and condemns her for it, through a criminal conviction, if she is proved guilty. One can of course count a criminal conviction as a kind of punishment: but it does not entail the kind of materially burdensome punishment, imposed after conviction, with which penal theorists are primarily concerned.

Another possible motivation for the abolitionist objection to the concept of crime is a kind of moral relativism that objects to the ‘imposition’ of values on those who might not share them (Bianchi 1994: 71–97): but since abolitionists are very ready to tell us, insistently, how we ought to respond to conflicts or troubles, and how a state ought or ought not to treat
its citizens, such an appeal to relativism reflects serious confusion (see Williams 1976: 34–39). More plausibly, the abolitionist claim could be that rather than take wrongdoing as our focus, we should focus on the harm that has been done, and on how it can be repaired; we will return to this suggestion in s.7 below.

Another abolitionist concern is that by defining and treating conduct as ‘criminal’, the law ‘steals’ the conflicts which crime involves from those to whom they properly belong (Christie 1977): instead of allowing, and helping, those who find themselves in conflict to resolve their trouble, the law takes the matter over and translates it into the professionalised context of the criminal justice system, in which neither ‘victim’ nor ‘offender’ is allowed any appropriate or productive role. Now it is a familiar and disturbing truth that our existing criminal processes — both in their structure and in their actual operations — tend to preclude any effective participation by either victims or offenders, although an adequate response to the criminal wrong that was done should surely involve them both. One response is to argue, as some abolitionists do, that our response to crime should consist not in punishment, but in a process of mediation or ‘restoration’ between victim and offender (see further s. 7 below); but another is to insist that we should retain a distinctive criminal process of trials, and punishment, in which the polity as a whole, acting on behalf of the victim as well as on its own behalf, calls the criminal wrongdoer to account — but that victims and offenders should be given a more active role in that process (see further Duff et al 2007, esp. chs. 3–5, 7). Such an insistence on the need for a public criminal process reflects two aspects of the concept of crime: first, it is sometimes important to recognise that a situation involves not just people in ‘conflict’, but a victim who has been wronged and an offender who has done the wrong; second, some such wrongs are ‘public’ wrongs in the sense sketched above — wrongs that properly concern not just those directly affected, but all members of the political community. Faced, for instance, by feuding neighbours who persistently accuse each other of more or less trivial wrongs, it might indeed be appropriate to suggest that they should forget about condemning each other and look for a way of resolving their conflict. But faced by a rapist and the person he raped, or by a violent husband and the wife he has been beating up, it would a betrayal both of the victim and of the values to which we are supposedly committed to portray the situation merely as a ‘conflict’ which the parties should seek to resolve: whatever else or more we can do, we must recognise and declare that here is a victim who has been seriously wronged; and we must be collectively ready to censure the offender’s action as a wrong (for a useful discussion of the significance of criminal law in the context of domestic violence, see Dempsey 2009).

However, to argue that we should retain the concept of crime, that we should maintain a criminal law which defines and condemns a category of ‘public’ wrongs, is not yet to say that we should maintain a penal system that punishes those who commit such wrongs; while a system of criminal law might require something like a system of criminal trials that will authoritatively identify and condemn criminal wrongdoers, it does not of its nature require the imposition of further sanctions on such wrongdoers. So we must turn now to the question of what could justify such a system of punishment.

3. Consequentialist Accounts

Many people, including those who do not take a consequentialist view of other matters, think that any adequate justification of punishment must be basically consequentialist. For we have here a practice that inflicts, indeed seeks to inflict, significant hardship or burdens: how else could we hope to justify it than by showing that it brings consequential benefits sufficiently large to outweigh, and thus to justify, those burdens? We need not be Benthamite utilitarians to be moved by Bentham’s famous remark that “all punishment in itself is evil. ... [I]t ought at all to be admitted, it ought only to be admitted in as far as it promises to exclude some greater evil” (Bentham 1789: ch. XIII.2). However, when we try to flesh out that simple consequentialist thought into something closer to a full normative account of punishment, problems begin to appear.

A consequentialist must justify punishment (if she is to justify it at all) as a cost-effective means to certain independently identifiable goods (for two simple examples of such theories, see Wilson 1983; Walker 1991). Whatever account she gives of the final good or goods at which all action ultimately aims, the most plausible immediate good that a system of punishment can bring is the reduction of crime. A rational consequentialist system of law will define as criminal only conduct that is in some way harmful; in reducing crime we will thus be reducing the harms that crime causes. It is commonly suggested that punishment can help to reduce crime by deterring, incapacitating, or reforming potential offenders (though for an argument that incapacitation is not a genuinely punitive aim, see Hoskins 2016: 260). (There are of course other goods that a system of punishment can bring. It can reassure those who fear crime that the state is taking steps to protect them — though this is a good that, in a well-informed society, will be achieved only insofar as the more immediate preventive goods are achieved. It can also bring satisfaction to those who want to see wrongdoers suffer — though to show that to be a genuine good, rather than merely a means of averting vigilantism and private revenge, we would need to show that it involves something more than mere vengeance, which would be to make sense of some version of retributivism.)

It is a contingent question whether punishment can be an efficient method of reducing crime in any of these ways, and some objections to punishment rest on the empirical claim that it cannot be — that there are other and more efficient methods of crime reduction (see Wootton 1963; Menninger 1968; Boonin 2008: 53, 264-67). Our focus here, however, will be on the moral objections to consequentialist accounts of punishment — objections, basically, that crime-reductive efficiency does not suffice to justify a system of punishment.

The most familiar line of objection to consequentialist penal theories contends that consequentialists would be committed to regarding manifestly unjust punishments (the punishment of those known to be innocent, for instance, or excessively harsh punishment of the guilty) to be in principle justified if they would efficiently serve the aim of crime
reduction: but such punishments would be wrong, because they would be unjust (see e.g., McCloskey 1957: 468-69; Hart 1968, chs. 1–2; Ten 1987; Primoratz 1999, chs. 2–3; Boonin 2008: ch. 2).

There are some equally familiar consequentialist responses to this familiar objection. One is to argue that such ‘unjust’ punishments would be justified if they would really produce the best consequences (see e.g., Smart 1973: 69–72; Bagaric and Amarasekara 2000) — to which the critic will reply that we cannot thus put aside the moral significance of injustice. Another is to argue that in the real world it is extremely unlikely that such punishments would ever be for the best, and even less likely that the agents involved could be trusted reliably to pick out those rare cases in which they would be: thus we, and especially our penal officials, will do best if we think and act as if such punishments are intrinsically wrong and unjustifiable (see e.g., Rawls 1955; Hare 1981, chs. 3, 9.7) — to which the critic will respond that this still makes the wrongness of punishing a known innocent contingent on its effects, and fails to recognise the intrinsic wrong that such punishment does (see e.g., Duff 1986: 151–64; Primoratz 1999, chs. 3.3, 6.5). Another response is to argue that a richer or subtler account of the ends that the criminal law should serve will generate suitable protection against unjust punishments (see Braithwaite and Pettit 1990, especially 71–76, on ‘dominion’ as the end of criminal law); but the objection remains that any purely consequentialist account will make the protection of the innocent against injustice contingent on its instrumental contribution to the system’s aims (on Braithwaite and Pettit, see von Hirsch and Ashworth 1992; Duff 1996: 20–25; Pettit 1997).

4. Retributivist Accounts

Whereas consequentialist accounts regard punishment as justified instrumentally, as a means to achieving some valuable goal (typically crime reduction), retributivist accounts contend that punishment is justified as an intrinsically appropriate, because deserved, response to wrongdoing (but see Berman 2011 for an argument that some recent versions of retributivism actually turn it into a consequentialist theory).

Theorists have distinguished ‘positive’ and ‘negative’ forms of retributivism. Positive retributivism holds that an offender’s desert provides a reason in favor of punishment; essentially, the state should punish those found guilty of criminal offenses to the extent that they deserve, because they deserve it. Penal desert constitutes not just a necessary, but an in-principle sufficient reason for punishment (only in principle, however, since there are very good reasons — to do with the costs, both material and moral, of punishment — why we should not even try to punish all the guilty). Negative retributivism, by contrast, provides not a positive reason to punish, but rather a constraint on punishment: punishment should be imposed only on those who deserve it, and only in proportion with their desert. Because negative retributivism represents only a constraining principle, not a positive reason to punish, it has been employed in various mixed accounts of punishment, which endorse punishment for consequentialist reasons but only insofar as the punishment is no more than is deserved (see s. 6 below).

A striking feature of penal theorising during the last three decades of the twentieth century was a revival of positive retributivism — of the idea that the positive justification of punishment is to be found in its intrinsic character as a deserved response to crime (see H. Morris 1968; N. Morris 1974; Murphy 1973; von Hirsch 1976; two useful collections of contemporary papers on retributivism are White 2011 and Tonty 2012).

Positive retributivism comes in very different forms (Cottingham 1979). All can be understood, however, as attempting to answer the two central questions faced by any retributivist theory of punishment. First, what is the justificatory relationship between crime and punishment that the idea of desert is supposed to capture: why do the guilty ‘deserve to suffer’ (see L. Davis 1972) — and what do they deserve to suffer (see Ardal 1984; Honderich 2005, ch. 2)? Second, even if they deserve to suffer, or to be burdened in some distinctive way, why should it be for the state to inflict that suffering or that burden on them through a system of criminal punishment (Murphy 1985; Husak 1992; Shafer-Landau 1996; Wellman 2009)?

One retributivist answer to these questions is that crime involves taking an unfair advantage over the law-abiding, and that punishment removes that unfair advantage. The criminal law benefits all citizens by protecting them from certain kinds of harm: but this benefit depends upon citizens accepting the burden of self-restraint involved in obeying the law. The criminal takes the benefit of the self-restraint of others, but refuses to accept that burden herself: she has gained an unfair advantage, which punishment removes by imposing some additional burden on her (see H. Morris 1968; Murphy 1973; Sadurski 1985; Sher 1987, ch. 5; Adler 1992, chs. 5–8; Dagger 1993, 2003, 2011; Stichter 2010; for criticism, see Burgh 1982; Duff 1986, ch. 8; Falls 1987; Dolinko 1991; Anderson 1997; Boonin 2008: 119–143; Hoskins 2011b).

This kind of account does indeed answer the two questions noted above. What the criminal deserves to suffer is the loss of her unfair advantage, and she deserves that because it is unfair that she should get away with taking the benefits of the law without accepting the burdens on which those benefits depend; it is the state’s job to inflict this suffering on her, because it is the author or guarantor of the criminal law. However, such accounts have internal difficulties: for instance, how are we to determine how great was the unfair advantage gained by a crime; how far are such measurements of unfair advantage likely to correlate with our judgements of the seriousness of crimes? (For a detailed defence of the ‘unfair advantage’ theory as a theory of sentencing, see M. Davis 1992, 1996; for criticism, see Scheid 1990, 1995; von Hirsch 1990.) Furthermore, they seem to misrepresent what it is about crime that makes it deserving of punishment: what makes murder, or rape, or theft, or assault a criminal wrong, deserving of punishment, is surely the wrongful harm that it does to the individual victim — not (as on this kind of account) the supposed unfair advantage that the criminal takes over all those who obey the law (for recent attempts to defend fair play retributivism...
A different retributivist account appeals not to the abstract notion of unfair advantage, but to our (normal, appropriate) emotional responses to crime: for instance, to the resentment or ‘retributive hatred’, involving a desire to make the wrongdoer suffer, that crime may arouse (see Murphy and Hampton 1988, chs. 1, 3); or to the guilt, involving the judgement that I ought to be punished, that my own wrongdoing would arouse in me (see Moore 1997, ch. 4). Such accounts try to answer the first of the two questions noted above: crime deserves punishment in the sense that it makes appropriate certain emotions (resentment, guilt) which are satisfied by or expressed in punishment. They do not yet show, however, why it should be the state’s task to satisfy or provide formal expression for such emotions (but see Stephen 1873: 152); and their answers to the first question are also problematic. Criminal wrongdoing should, we can agree, provoke certain kinds of emotion, such as self-directed guilt and other-directed indignation; and such emotions might typically involve a desire to make those at whom they are directed suffer. But just as we can agree that anger is an appropriate response to wrongs done to me, while also arguing that we should resist the desire to hit back that anger often, even typically, involves (see Horder 1992:194–7), so we could argue that although guilt, resentment and indignation are appropriate responses to our own and others’ wrongdoing, we should resist the desire for suffering that they so often involve. At the least we need to know more than we are told by these accounts about just what wrongdoers deserve to suffer, and why the infliction of suffering should be an appropriate way to express such proper emotions. (For critical discussions of Murphy, see Murphy and Hampton 1988, ch. 2; Duff 1996: 29–31; Murphy 1999. On Moore, see Dolinko 1991: 555–9; Knowles 1993; Murphy 1999. See also Murphy 2003, 2012.)

A third version of retributivism holds that when people commit a crime, they thereby incur a moral debt to their victims, and punishment is deserved as a way to pay this debt (McDermott 2001). This moral debt differs from the material debt that an offender may incur, and thus payment of the material debt (returning stolen money or property, etc.) does not settle the moral debt: punishment is needed to pay the moral debt, by denying the ill-gotten moral good to the perpetrator. Among the challenges for this account are to explain the nature of the moral good, how the offender takes this moral good from the victim, how punishment denies this good to the offender, and how doing so thereby pays the offender’s debt to the victim.

5. Punishment as Communication

Perhaps the most influential version of retributivism in recent decades seeks the meaning and justification of punishment as a deserved response to crime in its expressive or communicative character. (On the expressed dimension of punishment, see generally Feinberg 1970, Primoratz 1989; for critical discussion, see Hart 1963: 60–69; Skillen 1980; M. Davis 1996: 169–81.) Consequentialists can of course portray punishment as useful partly in virtue of its expressive character (see Lacey 1988; Braithwaite and Pettit 1990); but a portrayal of punishment as a mode of deserved moral communication has been central to many recent versions of retributivism.

The central meaning and purpose of punishment, on such accounts, is to communicate to offenders the censure or condemnation that they deserve for their crimes. Once we recognise, as we should, that punishment can serve this communicative purpose, we can see how such accounts begin to answer the two questions that retributivists face. First, there is an obviously intelligible justificatory relationship between wrongdoing and censure — as a response which is intended to impose a burden (the burden of condemnation by one’s fellows) on an offender for his offence: whatever puzzles there might be about other attempts to explain the idea of penal desert, the idea that wrongdoers deserve to suffer censure is surely unpuzzling. Second, it is appropriate for the state to ensure that such censure is formally administered through the criminal justice system: if crimes are public wrongs, breaches of the political community’s authoritative code, then they merit public censure by the community. Furthermore, although internal to censure is the intention, or hope, that the person censured will accept the censure as justified and will thus be motivated to avoid crime in future, this kind of account can avoid the charge (as brought against consequentialist theories) that it seeks to coerce or manipulate offenders into obeying the law. For censure addresses, and respects, the person censured as a rational and responsible agent: it constitutes an appropriate, deserved response to the wrong that she did, and seeks to bring her to modify her future conduct only by reminding her of the good moral reasons that she has for refraining from crime; it is an appropriate way for citizens to treat and respond to each other. (For different kinds of communicative account, see especially von Hirsch 1993, ch.2; Duff 2001, chs. 1.4.4, 3.2; Bennett 2008; Markel 2011, 2012. For critical discussion, see Davis 1991; Boonin 2008: 171–80; Hanna 2008; Matravers 2011).

However, an obvious and crucial question faces any such justification of punishment as a communicative enterprise. Censure can be communicated through a formal conviction in a criminal court; or it could be communicated by some further formal denunciation issued by a judge or some other representative of the legal community, or by a system of purely symbolic punishments which were burdensome only in virtue of their censorial meaning. It can, of course, also be communicated by ‘hard treatment’ punishments of the kinds imposed by our courts — by imprisonment, by compulsory community service, by fines and the like, which are burdensome independently of their censorial meaning (on ‘hard treatment’, see Feinberg 1970): but why should we choose such methods of communication, rather than methods that do not involve hard treatment (see Christie 1981: 98–105)? Is it because they will make the communication more effective (see Falls 1987; Primoratz 1989; Kleinig 1991)? But why is it so important to make the communication effective — and is there not a serious danger that the hard treatment will conceal, rather than highlight, the moral censure it should communicate (see Mathiesen 1990: 58–73)?

One sort of answer to this question explains penal hard treatment as an essential aspect of the enterprise of moral...
constraints might also be placed on the severity and modes of punishment that can be permitted: constraints either
expected to produce sufficient consequential benefits to outweigh its undoubted costs. (Further nonconsequentialist
imposing punishment on those who are eligible for it or liable to it: we should punish if and because this can be
permissible by the commission of the offence. The second step is then to offer positive consequentialist reasons for
the kinds of coercive treatment that punishment involves: such treatment, which is normally inconsistent with the
future choices by incapacitating them. And although a deterrent system does, unlike the others, offer potential offenders
it free, as responsible agents should be left free, to determine their own future conduct, but seeks to preempt their
reformative system treats those subjected to it not as rational, self-determining agents, but as objects to be re-formed
agent, we must seek to modify her conduct only by offering her good and relevant reasons to modify it for herself.
A simple version of this approach identifies certain side-constraints to which our pursuit of the consequential
benefits of punishment must be subject: constraints that forbid, for instance, the deliberate punishment of the innocent,
or the excessively harsh punishment of the guilty. (See most famously Hart 1968, and Scheid 1997 for a sophisticated
Hartian theory; on Hart, see Lacey 1988: 46–56; Morison 1988; Primoratz 1999, ch. 6.6.) Critics have charged that this
strategy is ad hoc or internally inconsistent (see Kaufman 2008: 45–49). In addition, retributivists argue that it relegates
retributivism to a merely subsidiary role, as a basis for side-constraints, when in fact giving offenders their just deserts
is a (or the) central rationale for punishment (see Wood 2002: 303).

6. Mixed Accounts

Given the challenges faced by pure consequentialist and pure retributivist accounts, some theorists have sought to
make progress on the question of punishment’s justification by incorporating consequentialist and nonconsequentialist
elements into their accounts. Perhaps the most influential example of a mixed account begins by recognizing that the
question of punishment’s justification is in fact several different questions, which may be answered by appeal to
different considerations: we can argue, first, that the ‘general justifying aim’ (Hart 1968: 8–11) of a system of
punishment must lie in its beneficial effects, but second, that our pursuit of that aim must be constrained by
nonconsequentialist principles that preclude the kinds of injustice alleged to flow from a purely consequentialist
account. A simple version of this approach identifies certain side-constraints to which our pursuit of the consequential
benefits of punishment must be subject: constraints that forbid, for instance, the deliberate punishment of the innocent,
or the excessively harsh punishment of the guilty. (See most famously Hart 1968, and Scheid 1997 for a sophisticated
Hartian theory; on Hart, see Lacey 1988: 46–56; Morison 1988; Primoratz 1999, ch. 6.6.) Critics have charged that this
strategy is ad hoc or internally inconsistent (see Kaufman 2008: 45–49). In addition, retributivists argue that it relegates
retributivism to a merely subsidiary role, as a basis for side-constraints, when in fact giving offenders their just deserts
is a (or the) central rationale for punishment (see Wood 2002: 303).

Another key worry about such accounts concerns the grounding of these side-constraints. If they are derived from a
‘negative’ retributivism that insists that punishment is justified only if it is deserved (see Dolinko 1991: 539–43), then
they face the thorny problem of explaining this retributivist notion of desert (see s. 4 above): but it is not clear whether
they can be justified without such an appeal to retributivist desert (see Hart 1968: 44–48; Feinberg 1988: 144–55;
Walker 1991, ch. 11). Even if such side-constraints can be securely grounded, however, consequentialist theories of
punishment face further, broadly Kantian objections, focused on the moral character of punishment within those
constraints. On such accounts, so long as punishment is deserved it may and should be used to serve consequentialist
ends — most obviously the end of crime reduction. But, the critic now objects, to use punishment thus is to use those
who are punished ‘merely as means’ to those further ends, which is to deny them the respect, the moral standing, that
is their due as responsible agents (see Murphy 1973: 218).

The Kantian prohibition on treating each other ‘merely as means’ is admittedly unclear in its implications (for a
useful discussion of how we should understand ‘the means principle’, see Tadros 2011: ch. 6). It can be argued that if
punishment is reserved for those who voluntarily break the law, it does not treat them *merely* as means (see Walker
1980: 80–85; Hoskins 2011a). Indeed, Kant himself suggested that as long as we reserve punishment only for those
found guilty of crimes, then it is permissible to punish with an eye toward potential benefits (Kant 1797: 473). Still, a
critic may argue that if we are to treat another ‘as an end’, with the respect due to her as a rational and responsible
agent, we must seek to modify her conduct only by offering her good and relevant reasons to modify it for herself.
Punishment aimed at deterrence, incapacitation, or offender reform, however, does not satisfy that demand. A
reformative system treats those subjected to it not as rational, self-determining agents, but as objects to be re-formed
by whatever efficient (and humane) techniques we can find. An incapacitative system does not leave those subjected to
it free, as responsible agents should be left free, to determine their own future conduct, but seeks to preempt their
future choices by incapacitating them. And although a deterrent system does, unlike the others, offer potential offenders
reason to obey the law, it offers them the wrong kind of reason: instead of addressing them as responsible moral
agents, in terms of the moral reasons which justify the law’s demands on them, it addresses them as merely self-
interested beings, in the coercive language of threat; deterrence treats ‘a man like a dog instead of with the freedom
and respect due to him as a man’ (Hegel 1821: 246. For these objections, see Lewis 1953; H Morris 1968; Duff 1986:

One strategy for dealing with them is to posit a two-step justification of punishment. The first step, which typically
appeals to nonconsequentialist values, shows how the commission of a crime renders the offender eligible for, or liable
to, the kinds of coercive treatment that punishment involves: such treatment, which is normally inconsistent with the
respect due to us as rational agents or as citizens, and inconsistent with the Kantian means principle, is rendered
permissible by the commission of the offence. The second step is then to offer positive consequentialist reasons for
imposing punishment on those who are eligible for it or liable to it: we should punish if and because this can be
expected to produce sufficient consequential benefits to outweigh its undoubted costs. (Further nonconsequentialist
constraints might also be placed on the severity and modes of punishment that can be permitted: constraints either
Thus, for instance, some argue that those who voluntarily break the law thereby forfeit at least some of the rights that citizens can normally claim: their wrongdoing therefore legitimises kinds of treatment (reformative or incapacitative treatment, for instance, or deterrent punishment) that would normally be wrong as violating citizens’ rights (see Goldman 1982; C Morris 1991; Wellman 2012; for criticisms, see Lippke 2001a; Boonin 2008: 103–19). We must ask, however, whether we should be so quick to exclude fellow citizens from the rights and status of citizenship, or whether we should not look for an account of punishment (if it is to be justified at all) on which punishment can still be claimed to treat those punished as full citizens. (The common practice of denying imprisoned offenders the right to vote while they are in prison, and perhaps even after they leave prison, is symbolically significant in this context: those who would argue that punishment should be consistent with recognised citizenship should also oppose such practices; see Lippke 2001b; Journal of Applied Philosophy 2005.)

Another view holds that punishment does not violate offenders’ rights insofar as they consent to their punishment (see Nino 1983). The consent view holds that when a person voluntarily commits a crime while knowing the consequences of doing so, she thereby consents to these consequences. This is not to say that she explicitly consents to being punished, but rather by her voluntary action she tacitly consents to be subject to what she knows are the consequences. Notice that, like the forfeiture view, the consent view is agnostic regarding the positive aim of punishment: it purports to tell us only that punishing the person does not wrong her, as she has effectively waived her right against such treatment. The consent view faces formidable objections, however. First, it appears unable to ground prohibitions on excessively harsh sentences: if such sentences are implemented, then anyone who subsequently violates the corresponding laws will have apparently tacitly consented to the punishment (Alexander 1986). A second objection is that most offenders do not in fact consent, even tacitly, to their sentences, because they are unaware either that their acts are subject to punishment or of the severity of the punishment to which they may be liable. For someone to have consented to be subject to certain consequences of an act, she must know of these consequences (see Boonin 2008: 161–64). A third objection is that, because tacit consent can be overridden by explicit denial of consent, it appears that explicitly nonconsenting offenders could not be justifiably punished on this view (ibid.: 164–165; but see Imbrisevic 2010).

Others offer contractualist or contractarian justifications of punishment, grounded in an account not of what treatment offenders have in fact tacitly consented to, but rather of what rational agents or reasonable citizens would endorse. The punishment of those who commit crimes is then, it is argued, rendered permissible by the fact that the offender himself would, as a rational agent or reasonable citizen, have consented to a system of law that provided for such punishments (see e.g., Dolovich 2004; Brettschneider 2007; Finkelstein 2011; for criticism, see Dagger 2011; see also Matravers 2000). Still others portray punishment (in particular deterrent punishment) as a species of societal (self-) defence — and it seems clear that to defend oneself against a wrongful attack is not to use the attacker ‘merely as a means’, or to fail to show him the respect that is his due. (For versions of this kind of argument, see Alexander 1980; Quinn 1985; Farrell 1985, 1995; Montague 1995; Ellis 2003 and 2012. For criticism, see Boonin 2008: 192–207. For a particularly intricate development of this line of thought, grounding the justification of punishment in the duties that we incur by committing wrongs, see Tadros 2011; for critical responses, see the special issue of Law and Philosophy, 2013.)

One might argue that the Hegelian objection to a system of deterrent punishment overstates the tension between the types of reasons, moral or prudential, that such a system may offer. Punishment may communicate both a prudential and a moral message to members of the community. Even before a crime is committed, the threat of punishment communicates societal condemnation of an offense. This moral message may help to dissuade potential offenders, but those who are unpersuaded by this moral message may still be prudentially deterred by the prospect of punishment. Similarly, those who actually do commit crimes may be dissuaded from reoffending by the moral censure conveyed by their punishment, or else by the prudential desire to avoid another round of hard treatment. What’s more, even if punishment itself did provide solely prudential reasons not to commit crimes, the criminal legal system more generally may communicate with citizens in moral terms. Through its criminal statutes, a community declares certain acts to be wrong and makes a moral appeal to community members to comply, whereas trials and convictions can communicate a message of deserved censure to the offender. Thus even if a system of deterrent punishment is itself regarded as communicating solely in prudential terms, it seems that the criminal law more generally can still communicate a moral message to those subject to it (see Hoskins 2011a).

A somewhat different attempt to accommodate prudential as well as moral reasons in an account of punishment begins with the retributivist notion that punishment is justified as a form of deserved censure, but then contends that we should communicate censure through penal hard treatment because this will give those who are insufficiently impressed by the moral appeal of censure prudential reason to refrain from crime; because, that is, the prospect of such punishment might deter those who are not susceptible to moral persuasion. (See Lipkin 1988, Baker 1992. For a sophisticated revision of this idea, which makes deterrence firmly secondary to censure, see von Hirsch 1993, ch. 2; Narayan 1993. For critical discussion, see Bottoms 1998; Duff 2001, ch. 3.3. For another subtle version of this kind of account, see Matravers 2000.) This kind of account differs from the accounts just discussed, on which retributivist prohibitions on punishment of the innocent or excessive punishment of the guilty constrain the pursuit of consequentialist aims, since in the current account the (retributivist) imposition of deserved censure is part of the positive justifying aim of punishment; and it can claim, in response to the Hegelian objection to deterrence, that it does not address potential offenders merely ‘like dogs’, since the law’s initial appeal to the citizen is in the appropriate moral terms: the prudential, coercive reasons flowing from an account of just what offenders render themselves liable to, or from other values external to the system of punishment.)
however, is that although we can learn much from the restorative justice movement, especially about the role that punishments such as Community Service Orders (now part of what is called Community Payback) or probation, it is the preferred mode of punishment in many penal systems, this suggestion will appear laughable; but if we think instead of restorative justice sometimes seek; but it could be apt to secure the kind of formal, ritualised reconciliation that is the secular penance, as sketched above. A system of criminal punishment, however improved it might be (see s. 8 below), was what fractured the relationship between offender and victim (and the broader community), and that is what must be recognised and 'repaired' or made up for if a genuine reconciliation is to be achieved. A restorative process that is to beappy must refer not only to whatever material harm was caused by the crime, but to the wrong that was done: that is, that they have done should not be done because it is morally wrong. Thus although the education view sets offender reform as an end, it also implies certain consequentialist constraints on how we may appropriately pursue this end. Another distinctive feature of the moral education view is that it conceives of punishment as aiming to confer a benefit on the offender: the benefit of moral education. Critics have objected to the moral education view on various grounds, however. Some are sceptical about whether punishment is the most effective means of moral education. Others deny that most offenders are in need of moral education; many offenders realise what they are doing is wrong but are weak-willed, impulsive, etc. Also, many liberal theorists object that the education view is inappropriately paternalistic in that it endorses coercively restricting offenders' liberties as a means to confer a benefit on them.

Each of the theories discussed in this section incorporates, in various ways, consequentialist and nonconsequentialist elements. Whether any of these is more plausible than pure consequentialist or pure retributivist alternatives is, not surprisingly, a matter of ongoing philosophical debate. One possibility, of course, is that none of the theories on offer is successful because punishment is, ultimately, unjustifiable. A question for those who endorse such a view is what alternative response to criminal offending we should endorse. The next section considers two candidate answers.

7. ‘Restorative Justice’ and Restitution

Abolitionist theorising about punishment takes many different forms, united only by the insistence that we should seek to abolish, rather than merely to reform, our practices of punishment. (Classic abolitionist texts include Christie 1977, 1981; Hulsmann 1986, 1991; de Haan 1990; Bianchi 1994; Golash 2005; and Boonin 2008.) In this section we will attend to just two types of abolitionist theory.

Many abolitionists look to ‘restorative justice’ as an alternative to punishment. (‘Restorative’ practices and programmes also play an increasingly significant, although still somewhat marginal, role within the criminal process of trial and punishment; but our concern here is with restorative justice as an alternative to punishment.) The restorative justice movement has been growing in strength: although there are different and conflicting conceptions of what ‘restorative justice’ means or involves, one central theme is that what crime makes necessary is a process of reparation or restoration between offender, victim, and other interested parties; and that this is achieved not through a criminal process of trial and punishment, but through mediation or reconciliation programmes that bring together the victim, offender and other interested parties to discuss what was done and how to deal with it (see generally Matthews 1988; Daly and Immageon 1998; von Hirsch and Ashworth 1998, ch. 7; Braithwaite 1999; Walgrave 2002; von Hirsch et al 2003; von Hirsch, Ashworth and Shearing 2005; London 2011; Johnstone 2011, 2012).

Advocates of restorative justice often contrast it with ‘retributive’ justice, and argue that we should look for restoration rather than retribution or punishment, and seek to repair harms caused rather than to inflict punitive suffering for wrongs done. But it could be argued that this is a mistake. For when we ask what it is that requires ‘restoration’ or repair, the answer must refer not only to whatever material harm was caused by the crime, but to the wrong that was done: that was what fractured the relationship between offender and victim (and the broader community), and that is what must be recognised and ‘repaired’ or made up for if a genuine reconciliation is to be achieved. A restorative process that is to be appropriate to crime must therefore be one that seeks an adequate recognition, by the offender and by others, of the wrong done—a recognition that must for the offender, if genuine, be repentant; and that seeks an appropriate apologetic reparation for that wrong from the offender. But those are also the aims of punishment as a species of secular penance, as sketched above. A system of criminal punishment, however improved it might be (see s. 8 below), is of course not well designed to bring about the kind of personal reconciliations and transformations that advocates of restorative justice sometimes seek; but it could be apt to secure the kind of formal, ritualised reconciliation that is the most that a liberal state should try to secure between its citizens. If we focus only on imprisonment, which is still often the preferred mode of punishment in many penal systems, this suggestion will appear laughable; but if we think instead of punishments such as Community Service Orders (now part of what is called Community Payback) or probation, it might seem more plausible.

This argument does not, of course, support that account of punishment against its critics. What it might suggest, however, is that although we can learn much from the restorative justice movement, especially about the role that
The previous sections sketched the central contemporary accounts of whether and how legal punishment can be justified—and some of the objections and difficulties that they face. A number of further important questions are relevant to theorising about punishment, which can only be noted here.

First, there are questions about sentencing. (On sentencing, see generally Robinson 1987; Morris and Tonry 1990; von Hirsch 1993; Tonry 1996; von Hirsch and Ashworth 2005; Ashworth, von Hirsch and Roberts 2009; Frase 2012.) Who should decide what kinds and what levels of sentence should be attached to different offences or kinds of offence: what should be the respective roles of legislatures, of sentencing councils or commissions, of appellate courts, of trial judges, of juries? By what criteria should such decisions be made: how far should they be guided by a retributivist principle of proportionality, requiring punishments to be ‘proportionate’ in their severity to the seriousness of the crime; how far by consequentialist considerations of efficient crime-prevention? What kinds of punishment should be available to sentencers, and how should they decide which mode of punishment is appropriate for the particular offence?

Second, there are questions about the relation between theory and practice — between the ideal, as portrayed by a normative theory of punishment, and the actualities of existing penal practice. Suppose we have come to believe, as a matter of normative theory, that a system of legal punishment could in principle be justified — that the abolitionist challenge can be met. It is, to put it mildly, unlikely that our normative theory of justified punishment will justify our existing penal institutions and practices: it is far more likely that such a theory will show our existing practices to be radically imperfect — that legal punishment as it is now imposed is far from meaning or achieving what it should mean or achieve if it is to be adequately justified (see Heffernan and Kleinig 2000). If our normative theorising is to be anything more than an empty intellectual exercise, if it is to engage with actual practice, we then face the question of what we can or should do about our current practices. The obvious answer is that we should strive so to reform them that they can be in practice justified, and that answer is certainly available to consequentialists, on the plausible assumption that maintaining our present practices, while also seeking their reform, is likely to do more good or less harm than abandoning them. But for retributivists who insist that punishment is justified only if it is just, and for communicative theorists who insist that punishment is just and justified only if it communicates an appropriate censure to those who deserve it, the matter is harder: for to maintain our present practices, even while seeking their radical reform, will be to maintain practices that perpetrate serious injustice (see Murphy 1973; Duff 2001, ch. 5).

Third, the relation between the ideal and the actual is especially problematic in the context of punishment partly because it involves the preconditions of just punishment. That is to say, what makes an actual system of punishment unjustified might be not its own operations as such (what punishment is or achieves within that system), but the absence of certain political, legal and moral conditions on which the whole system depends for its legitimacy (see Duff 2001, ch. 5.2). Recent scholarship on punishment has increasingly acknowledged that the justification of punishment depends on the justification of the criminal law more generally, and indeed the legitimacy of the state itself (see s. 2 above). For example, if the state passes laws criminalising conduct that is not justifiably prohibited, then this calls into question the justification of the punishment it imposes for violations of these laws. Questions of criminalisation are thus directly relevant to punishment’s justification (see Husak 2008). Similarly, if the procedures by which criminal justice officials apprehend, charge, and prosecute individuals are unjustified, then the subsequent infictions of punishment will be unjustified as well (see Ristrop 2015 and 2016; on specific aspects of criminal procedure, see, e.g., Loader 2014 on policing, Lippke 2011 on plea bargaining, Duff et al 2007 on criminal trials, Flanders 2013 on pardons). More broadly, if a states tolerates (or worse, encourages) serious social injustices, then this may undermine the state’s standing to punish offenders who are also victims of such injustices (see, e.g., Duff 2007, Holroyd 2010, Howard 2013).

Fourth, theorists of punishment should also attend to various kinds of coercive measure that can be imposed on those who have committed, or are thought likely to commit, crimes. These include the wide range of so-called ‘collateral’ consequences of criminal convictions: policies restricting offenders’ access to jobs, housing, public assistance, and a host of other goods; subjecting them to continued detention; making their criminal records publicly available, and so on.
Finally, theoretical discussions of criminal punishment and its justification typically focus, as this discussion has focused, on criminal punishment in the context of domestic criminal law. But a theory of punishment must also have something to say about its aims and justification in the context of international criminal law — about how we should understand, and whether and how we can justify, the punishments imposed by such tribunals as the International Criminal Court: for we cannot assume that a normative theory of domestic criminal punishment can simply be read across into the context of international criminal law (see Drumbl 2007). Rather, the imposition of punishment in the international context raises distinctive conceptual and normative issues. One key question is which crimes rise to the level of ‘international crimes’ and are thus rightly subject to prosecution and punishment by international rather than domestic institutions (for influential yet very different answers to this question, see Altman and Wellman 2004, May 2005, Luban 2010, and Duff 2010). Another important question is how international institutions should assign responsibility for crimes such as genocide, which are perpetrated by groups rather than by individuals acting alone. (Such questions arise in the domestic context as well, with respect to corporations, but the magnitude of crimes such as genocide makes the questions especially poignant at the international level.) Several scholars in recent years have suggested that rather than focusing only on prosecuting members of the groups responsible for mass atrocities, it may sometimes be preferable to punish the entire group qua group. A worry for such proposals is that they risk inflicting punitive burdens on innocent members of the group. In response to this concern, defenders of the idea of collective punishment have suggested that it need not distribute among the members of the group (see Erskine 2011, Pasternak 2011, Tanguay-Renaud 2013; but see Hoskins 2014b), or that the benefits of such punishment may be valuable enough to override concerns about harm to innocents (see Lang 2007: 255).

Bibliography

Primoratz 1999, Honderich 2005, and Ellis 2012 are useful introductory books. Duff and Garland 1994; Ashworth, von Hirsch; and Roberts 2009; and Tonry 2011 are useful collections of readings.


**Academic Tools**

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

**Other Internet Resources**

[Please contact the author with suggestions.]

**Related Entries**

criminal law, theories of | legal obligation and authority | legal reasoning: interpretation and coherence in | legal reasoning: precedent and analogy in | punishment

Copyright © 2017 by Antony Duff

Zachary Hoskins <Zachary.Hoskins@nottingham.ac.uk>
Introduction to Punishment. History. Legal Provisions on Treatment Instead of Punishment Approach in Austria, Poland and Spain. Resources. See Also. Modern punishments seek to incapacitate certain offenders who are believed to pose a threat to the community. Incapacitation involves restricting offenders' movement or liberty. In contemporary practice, the goal of incapacitation is associated with attempts to punish offenders in a manner proportionate to the seriousness of the crime.