1

Punishment as a Problem

The expression “the problem of punishment” is extraordinarily common; in fact, it has become a cliché in the specialized literature. Insofar as whatever else punishment seeks to do, it seeks to make wrongdoers suffer (by somehow diminishing their well-being or by visiting upon them something they do not want), it is immediately obvious that there is indeed something problematic about it, something in need of a justification. Understandably, much has been written about punishment and its justification. In spite of the ensuing voluminous and unwieldy literature, however, surprisingly little attention has been paid to the discussion of the type of problem – or indeed of the types of problems – that punishment generates.

These problems can usefully be broken down into two general types: practical and theoretical. When authors use the expression “the problem of punishment” (or its cognates), they above all have in mind the suffering brought about by the state (or other authorities). Throughout history states have often over-criminalized – that is, they have defined as punishable – far too many activities, and they have over-punished both those activities that should have been criminalized and those that should have not. But the extremes to which modern, conspicuously democratic states have taken over-criminalization and over-punishment are particularly worrying. Commenting on the United States, Douglas Husak notes: “in every 138

1 Compiling a comprehensive list of all uses of this expression is unnecessary. A few recent examples should suffice: “The Problem of Punishment” is the title of David Boonin’s 2008 book (Cambridge: Cambridge University Press), and it is the title of chapters or sections in books by J. Angelo Corlett (ch. 3 of his Responsibility and Punishment, Dordrecht: Springer [2013]), by Whitley R. P. Kaufman (ch. 1 of his Honor and Revenge: A Theory of Punishment, Dordrecht: Springer [2012]), and by Victor Tadros (first section of the first chapter in his The Ends of Harm: The Moral Foundation of the Criminal Law, Oxford: Oxford University Press [2011]).

2 For example, almost two millennia ago, commenting on the Lex Papia Poppaea, which criminalized (among other things) not marrying and not procreating, Tacitus complained: “[W]here the country once suffered from its vices, it was now in peril from its laws.” Further, reacting to the “countless and complex statutes” of his day, Tacitus added: “[W]hen the state was most corrupt, laws were most abundant.” See Tacitus, Annals (3.25) (John Jackson, trans.), Cambridge, MA: Harvard University Press (1931), 563.
residents is incarcerated. An estimated 1 in 20 children born in the United States is destined to serve time in a state or federal prison at some point in his life." The "grand total" of people under some form of supervision or another by the American criminal justice system is a staggering "over 7 million."

By and large, it has been these sorts of practical problems that have attracted the attention of contemporary punishment theorists. This focus may partly explain the fact that "the problem of punishment" is often taken to be either exclusively or predominantly "the problem of state punishment." There is no denying that these practical problems are both pressing and depressing. But this does not, I think, justify the comparatively little energy that has been expended in addressing the theoretical problem of punishment. In this book, thus, I attempt to shed much needed light on this theoretical problem.

I am moved by two convictions. First, I am convinced that the examination of theoretical problems, in general, is genuinely important in itself. For example, reflecting on his own views in “A Critique of Utilitarianism,” Bernard Williams remarked: “[T]hese considerations do not themselves give solutions to practical dilemmas . . . but I hope they help to provide other ways of thinking about them.” Williams’s view that utilitarianism is unable to cope with “the complexities of moral thought” has greatly advanced that general debate – even if it has not solved any practical problems. My approach in this book is very similar to, and in some ways inspired by, Williams’s, for I will defend the view that contemporary punishment theory is unable to cope with the complexity of moral thought and moral life. Thus, as a first approximation to my goals here, it can be said that I seek to deploy, within the specific context of punishment, criticisms similar to those more general criticisms Williams deployed against both utilitarianism and what he called “the morality system.” Second, I am convinced that the specialized literature on punishment has reached a kind of stalemate (which I will explain shortly). In light of this stalemate, the examination of the theoretical problems surrounding punishment may in fact help us better understand – and

4 I have criticized this view in Leo Zaibert, Punishment and Retribution, Aldershot: Ashgate (2006).
5 Bernard Williams “A Critique of Utilitarianism,” in J. J. C. Smart and B. A. O. Williams, Utilitarianism: For and Against, Cambridge: Cambridge University Press (1973), 117. Consider another eloquent example: “This book is unabashedly devoted to solving these problems, though to put it that way suggests an incredible hubris on the part of the author, and might also mislead the reader into thinking that the book is intended to put these problems, once and for all, to rest. It would be just as accurate to describe the book’s aim as to provide a way of understanding – or, if you like, interpreting – these problems,” in Susan Wolf, Freedom within Reason, Oxford: Oxford University Press (1990), 4.
6 Williams, “A Critique . . .,” 149.
eventually even help us solve – some of those practical problems that have hitherto monopolized attention. A full appreciation of the importance of the overlooked theoretical problem will, of course, develop slowly, as the book progresses. For now, an account of what the theoretical problem is shall suffice.

We have just seen the essential conceptual connection between punishment and suffering (understood very generally). The other essential conceptual connection worth our attention links punishment and (perceived) wrongdoing. The idea is that by making wrongdoers suffer, justice is achieved. Punishment is thus immediately revealed as generating the theoretical problem of having to bring justice through suffering. There is, I will assume, obvious value in diminishing suffering in the world and obvious value in imparting justice. By and large, I will focus on one specific type of suffering-diminution – variously called forgiveness, mercy, leniency (etc.) – and on one specific type of justice-imparting – punishment. It is obvious, too, that these values can conflict with each other, independently of any practical problems.

Theoretically speaking, then, punishment presents us with a moral dilemma: Which of these conflicting values is weightier? Importantly, and in contradistinction to the virtually universally accepted position, I will not assume that the dilemma simply evaporates, in the sense of being fully resolved, when punishment is taken to be (or not taken to be) justified. The expression “justification” (and cognates), particularly in the way contemporary punishment theorists use it, is far too impoverished to match the complexity of punishment and concomitant phenomena. Thus, I will here attempt to place the discussion of punishment in much closer proximity to discussions of moral dilemmas in general. I will in fact suggest that punishment theory has developed in remarkable isolation from other general advances in moral philosophy – in particular those associated with the budding specialized literature on moral dilemmas.

Punishment, I will argue, presents us with precisely the sorts of famous quandaries generated by other moral dilemmas – above all those associated with forgiveness and related phenomena. I will evidently have much more to say about forgiveness later on, but I would at once wish to suggest two important features of my understanding of this concept. First, the essence of forgiveness is the idea of a deliberate refusal to punish – it is a form of sparing (deserved) suffering that is motivated by a special moral reason. Second, while the idea of mercy is admittedly more general than the idea of forgiveness, I will treat both as synonyms here. Surely one could show mercy

7 For ease of exposition, I will henceforth ignore the “perceived” rider and assume that perceived wrongdoing is always correctly perceived wrongdoing.

8 I do not think that this assumption is problematic. Humans do tend to disvalue both gratuitous suffering and obvious injustice. I will not engage in this book with those thoroughgoing forms of skepticism, fatalism, incompatibilism, or determinism that call into question moral responsibility in general. Noting this assumption would scarcely be necessary were it not for the fact that there are influential authors who, notwithstanding their opposition to these thoroughgoing forms of skepticism, defend forms of limited skepticism aimed specifically at the moral defensibility (or metaphysical possibility) of deserved punishment.
absent any wrongdoing: the “mercy killing” of a patient afflicted with a painful terminal disease is evidently not a matter of forgiving her; helping someone in need may be an act of mercy, but, again, it is not a matter of forgiving her; etc. But when authors talk about mercy within the context of punishment theory, they are often talking about what I am here calling forgiveness: mercy in this context means the deliberate sparing of a suffering that is deserved as a result of wrongdoing (again: based on a special moral reason). The general, age-old tension between justice and mercy is, within the specific context of our reactions to wrongdoing, the tension between punishment and forgiveness.

In this chapter, I will begin to delineate the contours of a more comprehensive – and more complicated – approach to punishment. In the first section, I will place the (theoretical) problem of punishment within the context of general theodicies. One important goal of doing so is to highlight, as I do in Section 1.2, some central differences between axiological and deontic considerations. Although these differences are very well known in moral philosophy in general, they are typically overlooked by contemporary punishment theorists – to the detriment of that particular, specialized literature. In Section 1.3, I suggest a new way of understanding the central debate regarding the justification of punishment: instead of focusing on the distinction between retributivism and consequentialism, we should focus on the distinction between monistic and pluralistic justifications. The discussion of the differences between the axiological and the deontic on the one hand and of monism and pluralism on the other sets the stage for a general discussion, in Section 1.4, of the nature and structure of those moral dilemmas that I suggest are very similar to punishment and from whose consideration the specialized literature on punishment stands to benefit. With these initial pieces more or less in place – or at least in sight – I conclude the chapter with an overview of the remainder of the book.

1.1 PUNISHMENT, THEODICIES, AND MEANING

As plain as it is that punishment seeks to cause suffering, it is also plain – if not plainer – that punishment is not the only, or even the main, source of suffering in the world. The existence of suffering in general has always stood in need of an explanation. Helping us to overcome, to reduce, or at least to cope with suffering in general is an essential aspect of every major comprehensive worldview – from the most

secular to the most religious. G. W. Leibniz coined a very useful term to refer precisely to the systematic effort to explain why there is evil in the world: theodicy.\(^10\) In their theistic version, theodicies seek to explain why an allegedly benevolent and omnipotent God allows suffering to exist. But secular theodicies, as Max Weber’s towering work underscores, are also perfectly possible and illuminating.\(^11\) Human beings, independently of their religious commitments and independently of their differing degrees of intellectual sophistication, have naturally been attracted to theodicies since long before the term was coined. Trying to make sense of the fact of misery in the world is an essential part of the human condition – and this is what theodicies seek to do.

The central question of theodicy can be posed in a variety of ways, and in the previous paragraph I have in fact deliberately phrased it using three different terms – “suffering,” “evil,” and “misery” – precisely in order to highlight the irrelevance, for my purposes, of terminological minutiae. We well know the meaning of claims to the effect that punishment causes suffering, and we well know what it is to wonder why there is evil or misery in the world. Investigating the precise meaning of these terms may be an important and worthwhile project in some contexts, but not in ours. The fundamental question of theodicy can be expressed very generally: Why do bad things happen?

The expression “bad things” is, within the context of theodicies, necessarily – and unproblematically – loose, and it is indeed consistent with things that cause “suffering,” with “misery,” with things that can be considered “evil,” and with a host of other possible descriptions. Those who have felt the existential pull of the fundamental question of theodicy have not had, and have not needed, any precise definition of these terms. In fact, this unproblematic looseness regarding the “bad things” of interest to theodicies is conspicuous, too, in connection to the “bad things” that punishment is supposed to involve. The most influential contemporary definition of punishment, the Flew–Benn–Hart definition, clearly exhibits this looseness: in Hart’s own words, punishment “must involve pain or other consequences normally considered unpleasant.”\(^12\) These things “normally considered unpleasant” in the definition of punishment evidently are among the “bad things” of concern to theodicies.


\(^{11}\) Max Weber, Economy and Society: An Outline of Interpretive Sociology (Guenther Roth, ed.), Berkeley, CA: University of California Press (1978). See also Frederick Neuhouser’s Rousseau’s Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition, Oxford: Oxford University Press (2008), where he argues that Rousseau is best seen as presenting a secular theodicy (of amour-propre) that is “more palatable” than Hegel’s secular theodicy (4).

\(^{12}\) H. L. A. Hart, Punishment and Responsibility: Essays in the Philosophy of Law, 2nd edn., Oxford: Oxford University Press (2008), 4. While I have objected to aspects of the Flew–Benn–Hart definition of punishment, I find this looseness unobjectionable. My own definition of punishment recognizes that the punisher wishes something to happen to the punishee that would “somehow offset” the “bad thing” that she has done. See my Punishment and Retribution, 31 ff.
The fundamental question of theodicy can only be meaningfully asked on the basis of assuming the truth of another view whereby the world would be better if fewer bad things happened. Other things equal, a world with less suffering in it is better than one with more suffering in it. And then the connection between the theodicies and punishment theory begins to emerge, for punishment involves the deliberate creation of a bad thing – suffering. This confronts us with the particularly poignant question about punishment: while the fundamental question of theodicy in general inquires about what reason could be adduced for there being any suffering in the world (assuming, to repeat, that less suffering is better than more suffering), punishment theorists need to explain why it is sometimes good to deliberately choose to create more suffering in the world.

As soon as the connection emerges, however, a potential objection arises too. The objection is that punishment is essentially a matter not of inflicting suffering but of something else: a matter of denouncing some acts, educating society, preventing or reallocating harms, or defending ourselves or our societies. If this objection succeeded, the particular question of punishment would lose at least part of its poignancy – for what we would be deliberately choosing to do when we punish would be to cause not suffering but something else. Suffering would be merely a side effect (although a perfectly foreseeable side effect) of what we really choose to bring about: denunciation, education, prevention, defense, etc.

But the objection fails. It fails, firstly, because of the abusively stipulative determination of what exactly it is that we are choosing to do. If I know that my choosing to ϕ will cause you suffering (even if that is not my main or direct intention) – how compelling is it to say that by choosing to ϕ I am not, eo ipso, thereby choosing to cause you suffering? This is reminiscent of what Pascal mocked as the “grand method of directing the intention.”

But, secondly, the objection fails more fundamentally as well, as it evinces a misunderstanding of what punishment itself must – on pain of incoherence – mean. As Wittgenstein, among myriad others, has put it: it is perfectly “clear” that just as “reward must be something acceptable [or pleasant],” punishment must be “something unacceptable [or unpleasant].”

In an important article that will occupy my attention later on (in Chapter 5), John Tasioulas usefully traces the genealogy of a powerful response to this objection. Tasioulas cites Williams’s pithy version of the response:

The idea that traditional, painful, punishments are simply denunciations is incoherent, because it does not explain, without begging the question, why denunciations have to take the form of what Nietzsche identified as the constant of punishment, “the ceremony of pain.”

Whatever specific punishments turn out to be, to the extent that they remain forms of punishment, they will necessarily have to (seek to) make the wrongdoer suffer. The point here is in no way specially linked to denunciatory or communicative theories of punishment in particular. The point applies with equal force to any account of punishment in which the conceptual connection between punishment and suffering is rejected.

Denunciatory theories of punishment are but an example of what in his Theodicy Leibniz usefully called “medicinal” punishments – those that seek “to correct the criminal, or at least to provide an example for others.” Rehabilitative, educative, preventive, or defensive approaches to punishment are all equally susceptible to Williams’s point, as long as these lose sight of the fact that, when we punish, any of these (or other) goals are achieved by means of making wrongdoers suffer. Conceptually, inflicting this suffering is not optional: to refuse to inflict this suffering (whether or not this refusal entails abandoning other ulterior goals as well) is to thereby refuse to punish. To punish, then, is to (try to) inflict suffering (or pain or misery or a bad thing, etc.) on someone as a response to her wrongdoing. Punishment without trying to inflict suffering is like gifting an object without intending to transfer any right over the thing gifted or like feeding someone without intending to give her some nourishment. This is not to abuse any definition (in the sense of Hart’s famous protestation) – it is merely to use one.

Leaving the objection behind, then, we can return to the peculiar poignancy of the question regarding the suffering that punishment causes. There is plenty of suffering in the world: suffering arising from the inevitable clashes between human vulnerabilities and the brute forces of nature or from human malice and cruelty or...
from sheer accidents and bad luck. But to punish is to deliberately bring bad things about, and this may suggest that punishment is, in itself, a matter of making the world worse, not better. This is, in fact, exactly Jeremy Bentham’s extraordinarily influential position on punishment: “[P]unishment in itself is evil,” to which he immediately adds that therefore it should only be admitted (if at all) when “it promises to exclude some greater evil.” I will argue at length later (especially in Chapter 3) that the Benthamite position is as widespread as it is deficient.

Cases in which suffering is deliberately inflicted and that do not seem to make the world worse are evidently easy to imagine. To continue with the Leibnizian motif, just consider any medical intervention that causes some suffering but is likely to prevent much greater suffering. Or consider any case in which a person or a group of people is made to suffer in order to spare greater suffering to a larger group of people. These inflictions of suffering do not really make the world worse, since they are best described, once we broaden the lens through which we look at them, as overall diminutions of suffering. Many popular justifications of punishment – such as the already-mentioned Benthamite utilitarian justification and the rehabilitative, educational, preventive, or defensive justifications – can be seen as medicinal in Leibniz’s sense. Proponents of these justifications see the badness of the suffering that punishment brings about as compensated by that badness which, ex hypothesi, punishment prevents. And so, in their view, punishment actually makes the world better, not worse.

But some have argued that there is, somehow, something good, in itself, about the suffering that punishment inflicts, independently of whether it prevents greater suffering. Leibniz is one example; his use of the term “medicinal” in this context is a put-down: “[T]rue retributive justice,” he tells us, goes – because it ought to go – “beyond the medicinal.” The “harmony of things” that is essential to Leibniz’s theodicy demands “evil in the form of suffering.” If the medicinal justifications were the whole story, then punishment would not really contribute much to the problem of suffering, for it would be plain that the suffering it generates is just necessary to avoid greater suffering. We would simply need to ensure that our calculations are correct. The medicinal approach – so influential nowadays, as I will argue – reduces the real and deep moral and political problems associated with punishment to mere “technological” ones, to echo Isaiah Berlin’s insightful deployment of this term, itself reminiscent of Leibniz’s deployment of the term

21 Leibniz, Theodicy . . ., 425
22 Ibid.
Leibniz – along with myriad other thinkers – believes this is too reductionist because the suffering constitutive of punishment is sometimes intrinsically good, even if it does not contribute to reducing further suffering. That problem – i.e., how punishment can be in this way intrinsically good – is the serious, non-medicinal, non-technological problem worth our attention.

Often, however, theodicies – particularly some theistic theodicies – attempt to solve this problem in suspect ways, by ascribing inscrutable wisdom to God. So, for example, a common move in these theodicies – including, in a way, Leibniz’s own – is to suggest that God, being not only all-good but also all-knowing, knows that the suffering in the world that strikes us as excessive is in fact optimal. Things are taken to be interconnected in ways that mere finite beings cannot understand: attempts to “eliminate” this or that episode of suffering in the world would have produced even more suffering – and God, unlike us, knows this. Aside from their obscurantism, these approaches may perhaps be guilty of simply transferring the reductive approach: it is not humans that do the number crunching but God – it is just that God’s calculations are too complicated for us. God would understand how the medicine is indeed medicine, how the technology operates, but the reductionism could still be there, only hidden from us. Opponents of the reductionist approach would object to mere transfers: they would want to ensure that reductionism is clearly rejected.

Secular theodicies fare better, as they tend to emphasize that the problem of theodicy is the problem of meaningful existence; it is the problem of making sense of a world that, as Weber noted, is filled with “undeserved suffering, unpunished injustice, and incorrigible stupidity.” For obvious reasons, I will here ignore the problem of incorrigible stupidity; but the twin problems of undeserved suffering and of unpunished injustice are, just as obviously, central to my aims in this book.

It is important to underscore that according to Weber suffering is really a problem for secular theodicies only when it is not deserved, and injustices are more of a problem when they are not punished. These two problems are not merely calculative, medicinal, or technological: a world in which vice was commonly rewarded and virtue commonly punished would be problematic even if it could be shown that it contained the minimum amount of suffering possible. And it would be problematic, first and foremost, because this state of affairs would strike us as existentially meaningless. As Weber put it, “the need for an ethical interpretation of the ‘meaning’ of the distributions of fortunes among men increased with the growing rationality of conceptions of the world.”

This search for meaning in the world is particularly urgent in the face of what Weber called “the incongruity between destiny and merit.” Gershon Shafir’s insightful interpretation of what “destiny” and “merit” mean for Weber is illuminating. With Shafir, we can see that Weber understands “destiny” as the worldview according to which the distribution of suffering in the world is disordered, arbitrary, and random — and thereby meaningless. “Merit,” in contrast, is for Weber related to a worldview according to which the distribution of suffering in the world is somehow “ordered” — and thereby meaningful. With Weber and Shafir, we see that humans can create meaning in an otherwise meaningless world by “infusing it with a rationalized ethic” so as to make “merit and destiny coincide.”

And we can then see, too, the reason for the pejorative sense in which Leibniz uses the term “medicinal” to refer to some justifications of punishment. As Shafir reminds us, Weber also had (in this context) a negative view of the medicinal, for although medicine “is capable of diminishing suffering,” “the point of view of medicine” is “itself meaningless.” Needless to say, this is not supposed to be an attack on medicine as such or on any other technological mechanism that seeks to reduce suffering. Rather, it is both a criticism of the strategy of reducing the world of value to mere medicine, to mere technology, and a reminder of the relatively tenuous connection between medicinal strategies and the larger story concerning the relation between punishment and the meaning of life.

The fact that, on this view, distributions according to merit suffuse the world with meaning shall be very important in the remainder of the book. Admittedly, merit is not the only means of infusing meaning into the world – though it surely is one important such means. To merit something means to deserve something; one’s merit is one’s desert. That these expressions are indeed synonymous is brought out by the fact that in some languages there simply is no word for “desert” other than “merit,” in the sense that to translate the English proposition that “she deserves X” one would in those other languages have to say “she merits X.”

The importance of merit – reflected not only in Leibniz and Weber but also in many other thinkers who

26 Ibid.
28 Ibid., 521.
29 Ibid., 524.
30 In The Book of Job in Medieval Jewish Thought (Oxford: Oxford University Press [2004]), Robert Eisen suggests that prominent among the reasons why Job has remained so fascinating through the ages is that it deals with “one of humanity’s deepest and most vexing problems.” This problem is none other than “the suffering of the righteous” – i.e., the problem of undeserved suffering (ibid., 3). Job’s protestations are predicated on the fact that he (rightly) considers himself undeserving of suffering. In many interpretations (illuminatingly discussed by Eisen) Job comes to accept that God’s decision to inflict undeserved suffering upon the righteous in order to test their devotion is another means to infuse meaning into the world. One need not agree with this other means in order to see that Job is centrally concerned with finding meaning in suffering.
31 This is explicitly the case at least in Spanish and other Romance languages. Comparisons between “ordinary” ways of saying things in different languages reveals obvious — though overlooked — problems for versions of the “ordinary language” approach to philosophy. Not that the intimate